LIBRARY
JUL 31 1926

VOL. X JUNE, 1926 No. 2

The Journal of APPLIED PSYCHOLOGY

EDITORS

JAMES P. PORTER
Ohio University

WILLIAM F. BOOK
Indiana University

WITH A

BOARD OF CO-OPERATING EDITORS

TABLE OF CONTENTS

A. R. GILLILAND. A Revision and Some Results with the Moore-Gilliland Aggressiveness Test. 143
S. M. NEWHALL. Sex Differences in Handwriting 151
THOMAS C. BUECHARD AND CARL J. WARDEN. The Effect of the Size of the Advertising Section upon the Value of Individual Advertisements in It. 162
CLYDE W. GLEASON. An Experiment in Testing Engine Lathe Aptitude 171
JOHN J. B. MORGAN AND HAZEL LUCILLE HULL. The Measurement of Persistence 180
NELLIE E. POWERS. The Classification of Mental Deficients (Especially of Girls) for a Special Class Center 188
EVELYN EASTMAN. An Information Test Applied to Juvenile Delinquents 202
MYRTLE RAYMAKER WORTHINGTON. A Study of Some Commonly Used Performance Tests 216
GRACE ARTHUR. A Group Point Scale for the Measurement of Intelligence. A Further Study in Method 228
GLADYS C. SCHWESINGER. Slang as an Indication of Character. 245
EDITH M. PEYTON AND JAMES P. PORTER. Old and New Methods of Teaching Primary Reading 264
JAMES D. WEINLAND. The Perceptibility of Automobile License Numbers 277
NOTES AND NEWS 284
BOOK REVIEWS 289
NEW BOOKS AND PAMPHLETS RECEIVED 299

PUBLISHED BY THE EDITORS
THE WILLIAMS & WILKINS CO.
Mount Royal and Guilford Avenues
Baltimore, Md.

Entered as second class matter June 3, 1924, at the post office at Baltimore, Maryland, under the act of March 3, 1879.
Acceptance for mailing at special rate of postage provided for in section 1103, act of October 3, 1917 authorized December 21, 1922.

COPYRIGHT 1926, BY JAMES P. PORTER

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. V. Bingham, Carnegie Institute of Technology; June E. Downey, University of Wyoming; Robert H. Gault, Northwestern University; William Healy, Baker Foundation, Juvenile Court, Boston; H. L. Hollingworth, Columbia University; Joseph Jastrow, University of Wisconsin; Herbert S. Langfeld, Princeton University; James H. Leuba, Bryn Mawr College, Henry C. McComas, Princeton University; Walter D. Scott, Northwestern University; C. E. Seashore, University of Iowa; Daniel Starch, Harvard Graduate School of Business Administration; Edward K. Strong, Jr., Leland Stanford Jr. University; L. M. Terman, Leland Stanford Jr. University; E. L. Thorndike, Columbia University; J. E. Wallace Wallin, Teachers College Miami University; G. M. Whipple, University of Michigan; Helen Thompson Woolley, Merrill-Palmer School, Detroit.

The scope of The Journal of Applied Psychology will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeal. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environment conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or musical instruments, sports, etc.

Communications, manuscripts, reviews, summaries, notices, books and off-prints for review should be addressed to Dr. James P. Porter, Ohio University, Athens, Ohio. Contributors are requested to follow the conventional method of citation employed by this Journal; in every case the year, volume and pages should be specified.

The Journal of Applied Psychology is published quarterly. The subscription price is $5.00 a year. Single numbers $1.50. Remittances and business communications should be addressed to the Editors, James P. Porter, Ohio University, Athens, Ohio, or to William F. Book, Indiana University, Bloomington, Indiana.

A REVISION AND SOME RESULTS WITH THE MOORE-GILLILAND AGGRESSIVENESS TEST

A. R. GILLILAND

Northwestern University

The measurement of traits of personality is probably receiving more attention than any other single problem in psychology. The first of these traits to be measured was intelligence. But it was soon discovered that while intelligence is the most important single trait, it is not the only one. Within the last few years numerous attempts have been made to measure traits other than intelligence. One of the earliest of these was the measurement of aggressiveness or forcefulness.

The first form of the test for the measurement of this trait was described in an earlier volume of this journal.[1] Since its construction the test has been revised and this revision has been used with several groups of college students and with a group of salesmen. It is the purpose of this article to describe the revision of the test and to present some of the results of its use.

In the earlier form of the test there were six parts, two of these, the "electric shock" and the "snake" test have been discarded, and a new test has been added. The test now consists of the following parts: (1) the eye movement distraction test, (2) the staring test, (3) the handwriting test, (4) the word reaction test, and (5) the time reaction test. The revised test together with the method of giving and scoring is herewith described.

TESTS I AND II, EYE MOVEMENT AND DISTRACTION TEST

The eye movement test consists in counting the number of movements of the eyes of the subject who is instructed to gaze steadily while adding

[1] Moore, H. T., and Gilliland, A. R.: The measurement of aggressiveness. Jour. Applied Psychol., June, 1921, pp. 97–118.

a series of numbers as rapidly as possible. The adding is by increments of one. The subject is given a number between 30 and 50. For example, if 31 is the given number the subject adds 1 and gives the result 32, then he adds 2 to this sum and gives the result 34, then he adds 3 to this sum and gives the result 37, etc. The adding continues until nine numbers have been added which in the above example would give a final sum of 76. If the subject makes a mistake in any addition, adds the wrong number, or loses his place in the series, the observer states his problem correctly as for example, "41 and 5 are?" and repeats the problem every two or three seconds until a correct answer is obtained. The subject is then instructed by a nod or command to continue until the series is completed.

In giving the first two tests the subject is given a preliminary practice in addition without the counting of eye movements or timing. This practice is followed by an addition in which the subject is allowed to add without staring, but the time required for the addition is recorded. This is followed by two additions in which the subject stares the observer squarely in the eyes. In these additions both the number of eye movements and the time required for the additions are recorded. Slight movements of the eyes as in looking from the eyes to some other part of the face are not counted; only those movements which are overt and apparent. Blinking of the eyes is not counted unless there is a movement of the eyes during the blink. A last addition is made without the staring but the time of the adding is recorded.

For the normal adding time, average the time for the first and last timed additions; and for the time with distraction, average the second and third timed additions, that is, the times during which the subject was staring. Record the normal adding time, the time with distraction, and the number of eye movements in the second and third additions, In these tests the observer has the difficult task of keeping the time required to perform the additions, counting the number of eye movements, and correcting the errors in the adding. Therefore wherever possible, two observers should be used, one for timing and counting the eye movements, and the other to give the numbers with which the additions are to begin and to correct the errors of the subject during the adding.

TEST III, WRITING TEST

For this test the subject must be provided with a sheet of paper and pencil or pen. On the paper he is instructed to write his name and address. The directions are: "Please write your name and address." No mention is made of the speed of the writing. If the subject inquires concerning the rate, he is told to write at his usual rate. The observer records the time required for the writing without attracting any more attention to the timing than possible. The subject is then told to re-

write his name and address *just as rapidly as he possibly can.* Maximum speed is indicated not only by the wording of the directions but also by emphasis in repeating them. The observer again records the time.

TESTS IV AND V, WORD AND TIME ASSOCIATION TESTS

After a preliminary practice in word association with three or four words such as "book," "grass," "run," and "home," the subject is given the words "enterprize," "success," "company," "danger," and "death." The observer records the subject's word responses to each of the five words as well as the reaction time for each. The reaction time should be measured to fifths of a second but this can be done fairly accurately with the use of a good stop watch.

METHODS OF SCORING THE TESTS

The scoring method makes deductions from 100 points for each failure of the subject to meet the five types of requirement of the test. Corresponding to the relative importance of the different tests a maximum possible deduction of 40 points is allowed for test I for eye movements and 10 points for the distraction time of test II, and 25 points each for the writing and association tests. Stated more in detail the method of deductions for each type of failure is as follows. In the staring test 5 points are to be deducted for each movement of the eyes up to eight movements. Two points are to be deducted for each second of average time increase in excess of three seconds due to distraction. In the writing experiment ½ point is to be deducted for each per cent of increase in excess of 20 per cent of the normal time required to write at the most rapid rate. For example, if the normal rate is 20 seconds and the most rapid rate is 14 seconds, the 20 minus 14 divided by 20 equals 30 per cent increase. This is 10 per cent more than 20 per cent (the amount of increase in rate for an aggressive individual for which no deduction is to be made) for which 5 points deduction is to be made. In the word association test 2 points are to be deducted for each negative response to "enterprize," "success," and "company." Responses like "initiative," "push," "money," "power," "activity," "scheme," and "wealth" are to be considered positive responses for which no deductions are to be made. Responses like "failure," "enemy," and "fear" are to be considered negative responses for which deductions are to be made. One point is to be deducted for each colorless or doubtful response to each of these three words. Two points are to be deducted for each vivid personal type of response to "death" and "danger." Examples of such colorful suggestive responses are "enemy," "grave," "killed," "funeral" and "accident." The maximum deduction on the basis of word content for test IV is 10 points. For the time of response in

test V one point is to be deducted for every fifth of a second in excess of two seconds for each of the five words. The maximum deduction to be allowed for time of response is 15 points.

The score of any subject may be obtained by totaling his deductions and subtracting the total from 100. For example, subject A made two eye movements in test I. His distraction time was 4 seconds in test II. His writing time was 10 seconds for normal writing and 8 seconds for his most rapid rate in test III. His responses to the five words were "mystery," "business," "busy," "afterwards," and "scared" in test IV. His reaction times were $1\frac{1}{2}$, $1\frac{3}{5}$, $1\frac{1}{2}$, $2\frac{1}{5}$ and 1 second. The deductions for this subject were therefore for test I, $2 \times 5 = 10$; for test II, $(4 - 3) \times 2 = 2$; for test III, $(10 - 8) \times 10 = 20$ per cent, and therefore no deductions; test IV, $1 + 0 + 1 + 2 = 4$; for test V, $0 + 0 + 0 + 1 + 0 = 1$. The sum of the deductions is 17 and the subject's score is 83.

The test was given to 315 college students. The average score for the group was 70.8 (median 72 P.E. 12). Of this group 185 were men and 130 women. The average for the men was 70.08 (median 73 P.E. $9\frac{1}{2}$) and for the women it was 68.9 (median 71 P.E. $12\frac{1}{2}$). These results indicate there is very little difference between the aggressiveness of the college man and the college woman. This seems reasonable. The greater variability of the women is probably to be explained by the fact that less initiative is required on the part of many girls than boys in attending college. On the other hand, a girl who must make her own way or meet other handicaps has a more difficult problem to face than the boy in similar circumstances. In other words, the girl is likely to be either more or less protected than the boy.

In order to be sure that the test is not just another test of intelligence, the scores on the aggressiveness test were correlated with intelligence test scores for three different groups. The first of these was a group of 50 liberal arts students for whom scores on the Brown University and three parts of the American Council of Education tests of intelligence as well as aggressiveness scores were available. The Pearson coefficient of correlation between the sum of the intelligence tests and the aggressiveness scores was +0.008. The correlation between aggressiveness scores and scholarship standing for these same students was +0.024. The correlation between intelligence

and aggressiveness for 55 students in the school of speech was −0.027. The aggressiveness test and the Scott Mental Alertness test were given to 42 salesmen of classified advertising. The correlation between the scores of the two tests was +0.239 (P.E. 0.01). This correlation while not high is much higher than others. Whether this was due to the nature of the Scott test, to the character of the salesmen as a result of natural and artificial selection, to chance in such a small group, or to other causes, it is impossible to determine.

From these data it is evident that the test of aggressiveness does not measure the same thing that is measured by the ordinary intelligence test. Whatever it be that is measured,

TABLE 1

Showing the percentile distribution of scores for 315 college students

PERCENTILE	SCORE	PERCENTILE	SCORE
0	26	60	76
10	46	70	80
20	56	80	85
30	64	90	91
40	69	100	100
50	72		

we can be sure it is something different. It now becomes our problem to try to discover whether the attribute or attributes measured by the test are to be found in situations where we would expect aggressiveness to occur. Of course we cannot hope to find a situation where this trait will be found isolated from all others. But we can find situations where the other traits tend to remain constant. Data from two sources already mentioned were available for this part of the study. These sources were the class in public speaking and the group of salesmen.

It seems reasonable to assume that aggressiveness along with intelligence is an important factor in public speaking. The speaker is instructed to look his audience squarely in the eyes, speak and assume a forceful attitude, and the like.

We have the intelligence scores, the aggressiveness scores, the grades in private work in speaking, as well as the final semester grades in speech. This last grade is a combination of the private work, written tests and grades on speeches made before the class. It should furthermore be noted that the quality of the speeches as well as the ability to present them was included in the grades in speech. The correlation between the grades in the private speeches and intelligence scores was +0.028. The correlation between intelligence scores and semester grades in speech was −0.125. If the grades in speech are reliable measures, intelligence, at least within the range of the intelligence to be found in a college class, is not a factor in successful public speaking.

Neither does aggressiveness appear to be an important factor in grades on private lessons in speech although it is a greater factor than intelligence. The correlation between the aggressiveness scores and the private lessons in speech was +0.20 (P.E. 0.01). However the correlation between aggressiveness and semester grades in speech was +0.34 (P.E. 0.08). This is a significant correlation, especially since the correlation between grades in speech and intelligence was negative. These results can be interpreted as meaning that aggressiveness as measured by this test is a rather important factor in success in public speaking. It is probably as important a factor as intelligence, when measured by the average group intelligence test, is a factor in scholarship.

It also seems very likely that aggressiveness plays no small part in success in many phases of modern business, especially in certain kinds of salesmanship. In order to further verify or disprove the value of our test, it was given to 43 salesmen in the classified ad department of a large Chicago newspaper. Due to difficulties in competition with salesmen of a larger newspaper, these salesmen had much difficulty in selling their space. In fact it was due to the heavy turnover that the writer was asked to assist in the selection of salesmen for the company.

So far as school training was concerned, the group varied from college graduates to some who had not finished the gram-

mar school grades. The average score of the group on the Scott Mental Alertness test was 49.5 or about five points above the average college freshman. The average of the group on the aggressiveness test was 69.2 or just the college average.

But more important than these averages, is the degree to which high scores or low scores on the aggressiveness test indicates success or failure in selling advertising. Unfortunately there was no completely objective criterion with which to check these scores. The amount of space sold was not fair

TABLE 2

RANK	AGGRESSIVENESS SCORE	MENTAL ALERTNESS SCORE	RANK	AGGRESSIVENESS SCORE	MENTAL ALERTNESS SCORE
1. Ar	96	69	14. Cu	63	40
2. Wa	94	57	15. Ro	49	57
3. Br	78	57	16. La	58	71
4. Je	86	73	17. Ro	45	35
5. Pi	48	33	18. Ku	94	39
6. Bu	56	30	19. De	74	64
7. Cu	91	49	20. Re	94	38
8. St	68	32	21. Ha	62	54
9. Ru	79	53	22. Ip	41	41
10. Mc	81	54	23. Ma	91	73
11. Ke	64	50	24. Rl	51	40
12. Cr	96	70	25. Ha	81	43
13. Ad	98	56	26. Am	64	30

since some of the salesmen had been with the company many years and others only a few weeks. Experience is a large factor in such selling. Some men had good territory and others poor; some had more territory than others; some had preferred kinds of space to sell such as "automobiles for sale." These and other causes made the amount of space sold an unreliable index of sales ability. The best standard of comparison obtainable was the rating of the men by the sales manager and his two assistants.

By dividing the salesmen into an upper and a lower half on the basis of the average of the ranks given by the three managers, the general superiority of the upper group on the aggressiveness test is shown (see table 2). In assigning these ranks so far as possible factors other than selling ability were eliminated. Specific mention was made by the sales manager regarding Pi. (5) and St. (8) who were given high rankings and received low scores on the tests. When these men first applied for a position with the company they were refused. They persisted and were finally taken on trial. They did poor work for the first year but were persistent and have at last become excellent salesmen. Persistence rather than intelligence or aggressiveness, seems to be their chief virtue. Ma. (23) had been with the company only a few weeks when the ratings were made. Before the study was completed Ma. was standing fourth in a contest among the salesmen and the manager admitted that he had probably been rated entirely too low on first impressions.

Combining the rating of the three sales managers for the twenty-six salesmen, the correlation with the aggressiveness scores was +0.26 (P.E. 0.129). Although this correlation is not high, it is more than the correlation of +0.204 (P.E. 0.133) between this ranking and the scores of the Scott Mental Alertness test. By the method of partial correlation, reducing this factor of mental alertness scores to a constant the correlation between rank and aggressiveness is +0.23.

These tests go to show that both in the case of students in public speaking and the salesmen of classified advertising, the results of the tests correlate positively and significantly with measures or estimates of success in these fields. It is clear that aggressiveness is not the only factor in these situations. There are many others, some of which we may soon be able to isolate and measure. It may be admitted also that the test in its present form is not a perfect measure of the trait we have called aggressiveness. But it is clear that the results of the use of the test measure an important human factor or factors; it does have a practical value and it is worthy of further study and use.

SEX DIFFERENCES IN HANDWRITING

S. M. NEWHALL

Yale University

INTRODUCTION

There are at least four investigations relevant to the problem of sex differences in handwriting. Awramoff and Meumann independently obtained pressure curves of handwriting by means of a tambour beneath the writing surface.[1] On the basis of differences in pressure and speed, as well as in size, rhythm and letter-formation, both concluded, Awramoff somewhat tentatively, that there are masculine and feminine writing types. Only a few subjects of each sex were used and several of the criteria were reversed, i.e., criteria which for the one experimenter were of the one sex, for the other were of the other sex.

Binet's experiment preceded a similar one by Downey.[2] The subjects of both investigations were engaged in estimating the sex of the authors of previously obtained handwriting specimens rather than in writing specimens themselves. In both instances, the samples were 200 or about 200 envelope addresses, half or approximately half written by members of each sex, and presented in unknown order. In both instances the subjects were unskilled, except for two of Binet's who were

[1] Awramoff, D. Arbeit und Rhythmus. Philos. Stud., 1903, 18, 559, 561.

Meumann, E. Vorlesungen zur Einführung in die Experimentelle Pädagogik u. s. w., 1907, Bd. 2, 289, 290.

[2] Binet, A. Les révélations de l'écriture d'après un contrôle scientifique, 1906.

Downey, J. E., Judgments on the Sex of Handwriting. Psychol. Rev., 1910, 17, 205.

"experts." A summary of numerical results for unpracticed subjects follows:

	MEAN PER CENT OF CORRECT JUDGMENTS	RANGE	N
Subjects of Binet	69.8	65.9–73.0	10
Subjects of Downey	67.3	60.0–77.5	13

Both investigators concluded that it is possible, with a certain amount of error, to determine sex from handwriting and Downey specified correctness "in perhaps eighty cases out of a hundred." An analysis of results by the latter led, further, to the belief that a typical feminine hand is "colorless, conventional, neat, and usually small;" while a typical masculine hand is "bold or careless or experienced, above all, individual."

All of the experimenters report "inversions of sex signs," i.e., instances in which an individual of one sex exhibits writing which seems more characteristic of the other sex. The last experimenter suggests that these inversions may depend on such factors as amount of writing done, age, or professional requirement; the latter factor coupled with early training being perhaps also a basis for the alleged sex differences themselves. Another possible basis for sex differences is suggested by the view that men exceed women in motor capacity.

THE PRESENT EXPERIMENT

The chief concern of the present experiment is the practical question of the degree to which the "ordinary," unpracticed or non-"expert" individual can determine sex from handwriting; more simply, what is the probability of correctness of judgment by such an individual. This study is characterized by the fact that a sufficiently large number of subjects served to make the results of statistical interest.

Materials and procedure

The handwriting specimens were, as in previous studies, 200 mail addresses; and half of them were written by each sex and

without knowledge that they would be used experimentally. The writer of each sample was a different individual and the range of ages was, roughly, 20 to 60. All of the addresses were of the same street, city, and state. In a few cases the street number was different though in all except one it contained the same number of digits. Care was taken to control possible secondary criteria, like color, quality, texture or perfume of stationery. There was no intentional selection of the handwriting itself.

The addresses were cut from the envelopes excluding the names of the persons addressed, postage stamps and cancellation marks. By excluding the names, a suggestion-error possibly present in other investigations was eliminated. The addresses were gummed to 3-inch by 6-inch cards for convenience in examining. These cards were then arranged in chance order, stamped with numbers consecutively from 1 to 200, punched with holes at one end, and stacked on metal rods in such a way as to facilitate separate inspection in a fixed order.

Record sheets were prepared which provided for 200 judgments and the placing of any judgment in any one of four categories of confidence. The subject was requested to "look at the first sample and judge whether it was written by a man or a woman." He was asked to "put an M or F (masculine or feminine) on the proper blank after No. 1," and to "turn over the first sample." "Proceed in the same way with the second sample, and so on with the rest." "Do not turn back for comparisons." . . . . "After finishing, state if possible, any characteristics of the handwriting which seemed to be important in making your judgments." The subjects were told nothing concerning the relative numbers of masculine and feminine samples or whether or not all samples were by one sex alone.

The subjects were 92 in number, unpracticed, and unfamiliar with the handwriting of the authors of the samples. All except 13 judged the samples twice in sessions separated by not less than twenty-four hours nor more than three weeks. Subsequent study indicated no difference in the order of the results obtained with the longer as compared with the shorter intervals.

There were 61 men subjects being, with 5 exceptions, undergraduate students taking courses in psychology. The exceptions were 2 professors and 3 graduate students in psychology. Of the 31 women subjects, 20 were Normal students or teachers taking summer school courses in psychology and other subjects. The remaining 11 included 3 graduate students in psychology, 4 typists and secretaries, 2 library workers, 1 teacher, and 1 individual who gave occupation as "none." The age-distribution for all subjects ran from 17 to 60 with 73 per cent of the cases in the twenties. The data were obtained during the academic year 1924–1925.

Results

The results relating to degree of confidence in judgment are not reported here, but will be in connection with another investigation.

Tables 1 and 2 give the means of the percentages of correct judgments for all subjects. The essential difference in these tables is that the first differentiates the trials and the second the sex of the subjects. As indicated below table 1, the individual percentages, from which the means were computed, were figured in two ways: (1) on the basis of the number judged correctly out of the total of 100 samples of the given sex, and (2) on the basis of the number judged correctly out of all samples *judged* to be of the given sex. In an ideal chance situation the percentages of both sorts would have the value 50.

In table 1 the differences between the first and second trials show chance conformation with the $P.E._{diff.}$ and, further, the difference is not always in the same direction. It is, therefore, to be regarded as insignificant. In each case, the $P.E._{diff.}$ is less than 1 per cent. The means and the probability of correctness are, therefore, significantly above 50. Approximately 85 per cent of the subjects judged with more than chance correctness.

The means of the frequencies of the "masculine" and "feminine" judgments, *irrespective of correctness*, are not shown in the tables. It may be noted, however, that in each of the four

situations in tables 1 and 2 (i.e., first and second trials, masculine and feminine groups of subjects), the mean frequency of the "masculine" judgments is greater than the "feminine,"

TABLE 1

Percentages of correct judgments of sex from handwriting

TRIAL	Mc	Fc	m	$\frac{Mc}{M}100$	$\frac{Fc}{F}100$
1	57.7	57	57.34	57.4	57.7
2	59	56	57.5	57.6	57.1
P.E.$_{diff.}$	0.95	0.94	0.66	0.63	0.68
$\frac{D}{P.E._{diff.}}$	1.4	1.1	0.2	0.3	0.9

The figures in the first trial row are means from 92 subjects and in the second trial row from 79 subjects.

Mc = number of handwriting specimens judged correctly to be masculine, out of a possible 100 correct.

Fc = number of handwriting specimens judged correctly to be feminine, out of a possible 100 correct.

m = mean.

$\frac{Mc}{M}100$ = percentage of specimens judged correctly to be masculine out of all specimens judged masculine.

$\frac{Fc}{F}100$ = percentage of specimens judged correctly to be feminine out of all specimens judged feminine.

TABLE 2

Relative proficiency of the masculine and feminine judges

JUDGES	Mc	Fc	m	$\frac{Mc}{M}100$	$\frac{Fc}{F}100$
Masculine	57.8	56.1	56.94	57.2	56.8
Feminine	59.0	57.6	58.44	58.1	58.8
$\frac{D}{P.E._{diff.}}$	1.3	1.6	2.1	1.3	2.4
Chances in 100	81.0	86.0	92.0	81.0	94.0

the difference running from 0.9 to 5.3 per cent. This is in the same direction as Downey's "distinct preponderance of masculine judgments."[3] In the latter case the frequency of the "masculine" judgments was about 9 per cent greater than that of the feminine.

Table 2 compares the frequency of *correct* judgments by the masculine subjects with that by the feminine. The means for the feminine subjects are seen to be somewhat the greater, and, though not shown in the tables, this is also the result for both first and second trials taken separately.

The desirability of computing measures of reliability before stating conclusions concerning sex differences, especially when the observed differences are not large, need not be emphasized here.[4] The statistical chances in 100 that the differences in table 2 would occur in the same direction on subsequent occasions, i.e., that women really surpass men, range from 81 to 94.

TABLE 3

Inter-correlations

AGE OF SUBJECT, DURATION OF SESSION	AGE OF SUBJECT, CORRECTNESS OF JUDGMENT	DURATION OF SESSION, CORRECTNESS OF JUDGMENT
0.11	0.24	0.01

Table 3 exhibits correlations, based on the first trial, among age of subject, duration of session, and correctness of judgment; and shows the three coefficients to be low and positive. The correlation of 0.24 $\pm$0.07 between age and correctness seems reasonable in view of the fact that we should expect correctness to increase with experience. The lowness of correlation may be due in part to lack of perfect correspondence between age and experience, to the short range of ages of the subjects as related to the precision of differentiation in handwriting judgments, and to overlapping of the sex criteria. Partial correlation, eliminating duration, does not affect this coefficient.

[3] Op. cit., 209.

[4] Frasier, G. W. Jour. Educational Psychol., 1920, 11, 34-38.

If the correlation between age and correctness is performed separately for each sex the results are as follows: Men, -0.07 ± 0.09; women, 0.53 ± 0.09. The insignificant corrrelation for the men is possibly due to their very short range of ages, short both with reference to possible ranges and to the range of the women subjects. The women's age-distribution is extremely skewed to the right; but even after the distribution is made to approximate normality by eliminating the extreme cases, the correlation is still as high as 0.40 ± 0.11.

Professor Downey kindly supplied the approximate ages of 12 of her 13 subjects and here also a positive relation was discovered between age and correctness, the correlation being over 0.60. Except for two or three, these subjects were all women.

It may be observed in connection with table 2 that the mean age of our feminine subjects was about five years in excess of the masculine mean; and possibly a correlative excess in feminine experience might account for the seeming sex difference in ability to judge handwriting. The mean age of Downey's subjects was also several years in excess of our masculine mean. It seems possible, on the other hand, that sex differences themselves, in handwriting as well as in ability to judge it, may depend chiefly on differences in experience which are *characteristic* of the sexes.

But whether or not that is so, it is of course certain that ability to judge handwriting depends in some degree on experience. Whether special practice or familiarity with handwriting can readily produce a noticeable improvement in estimating sex from it, is a question which has not received much attention. Downey allowed one of her subjects to review the errors made in the first 100 of her judgments and observed a higher per cent correct in the last 100 judgments.[5] Between the first and second trials of our experiment, 12 men subjects selected at random performed a special exercise. The subject was given two sets of 50 handwriting specimens, each set arranged like the main set of 200 except that one set of 50 contained only masculine specimens and the other only feminine

[5] Op. cit., 207.

specimens. The respective sets were labelled and the subject hence knew which specimens were written by which sex. He

TABLE 4

Effect of interpolating between the first and second trials

SUBJECT	TRIAL 1	TRIAL 2	INCREASE	DECREASE
A study and comparison of identified specimens				
47	52	56.5	4.5	
48	51.5	58	6.5	
49	58	56.5		1.5
50	56	59	3	
51	57	60	3	
52	58.5	56.5		2
53	64	63		1
54	59.5	50.5		9
55	51.5	54.5	3	
56	52.5	55	2.5	
57	47.5	49	1.5	
58	64	64		
Mean.....	56	56.87	2.0	1.1
Results of an equal number of random subjects who did not perform the interpolated exercise				
7	55	59	4	
8	62.5	65	2.5	
9	50	48		2
10	56.5	55		1.5
11	54	54.5	0.5	
12	46.5	49	2.5	
13	58.5	53.5		5
14	59.5	62	2.5	
15	65.5	58.5		7
16	52.5	51		1.5
17	57	58.5	1.5	
18	57.5	53.5		4
Mean.....	56.2	55.62	1.1	1.8

The figures in the "trial 1" and "trial 2" columns are individual percentages of correct judgments and correspond to the group means (m) in tables 1 and 2.

FIG. 1. THE 20 SPECIMENS MOST FREQUENTLY JUDGED MASCULINE OR FEMININE

was requested to "compare the two sets of known samples with the aim of learning to differentiate between the writing of men and women in general;" and to "make brief notes of any clues or marks of difference that seem important." This special exercise was performed in two periods, one just after the first trial of the main experiment and the other just before the second trial. The subject was asked to take as much time as seemed profitable. His notes were available for reference during the second trial of the main experiment. Table 4 compares performance in the second trial of the main experiment with the first trial, both for the experimental group and a control group of 12 men subjects. Inspection of the upper and lower halves of the table shows that the interpolated training had no significant influence on the percentages of correct judgments. For that reason the data from these subjects were also embodied in tables 1 and 2. Special experience of this character, then, would not appear to be specially effective.

Certain varieties of handwriting seem to be generally *regarded* as masculine and others as feminine. The sex of the writer, moreover, does not necessarily correspond to the general judgment; in other words, there are "inversions of sex signs." Both of these points are illustrated in figure 1 which may be explained as follows: The frequency, in the first trial, with which each of the 200 specimens was judged to be masculine or feminine by the 92 subjects was determined. The 10 specimens most frequently *judged* masculine and the 10 most frequently judged feminine were then abstracted and arranged in order of frequency. The first line of each of these specimens is reproduced in figure 1, those judged masculine being on the left. Both dimensions are reduced about one-half. The numbers and letters on the central strip show the frequency with which the respective samples were judged masculine or feminine and whether they actually were written by men or women. At the top of the column, e.g., "82 M" means that 82 of the subjects judged sample 130 to be masculine and that it actually was masculine. At the bottom of the figure, "75 F" to the right of sample 113 means that 75 subjects judged that sample masculine, though it was actually feminine. Samples

113 and 105 are the only inversions which appear in the figure. In other words, 18 out of the 20 most frequent judgments were correct. In every instance where more than 75 of the judges, or 82 per cent agreed, the judgment was correct.

The subjects reported scores of criteria of masculine and feminine writing, but as the writer knows of no way of discovering which were actually used in judging and which were rationalizations, they are not reproduced here. Presentation of a photographic reproduction of the specimens themselves seemed the best way of showing such typical differences as might be observable. In these 20 extreme cases, one is struck by the angles, irregularity, and slant of the specimens judged masculine and by the curves, regularity, and verticality of the feminine. Most of the feminine specimens, also, seem to be somewhat smaller.

SUMMARY

1. According to our results, the probability of non-"experts" correctly estimating sex from handwriting is of the order 0.55 or 0.60. This probability is not sufficiently high for individual judgments to be of practical value. It may be recalled, however, that various secondary criteria which are frequently present in everyday life were eliminated in this investigation.

2. There is some evidence of a positive relation between age of subject and ability to estimate sex from handwriting. It seems possible that a selection of older subjects might show a probability of correctness comparable to the results of previous investigators.

3. There is somewhat doubtful evidence that women are superior to men in estimating sex from handwriting.

4. The handwriting most frequently judged to have been written by men appears characteristically different from that most frequently judged to have been written by women.

5. In every instance examined where over 82 per cent of the subjects agreed in their judgment of a handwriting, the judgment was correct. It seems possible that group judgments may be of practical value.

THE EFFECT OF THE SIZE OF THE ADVERTISING SECTION UPON THE VALUE OF INDIVIDUAL ADVERTISEMENTS IN IT

THOMAS C. BURCHARD AND CARL J. WARDEN

Psychological Laboratory, University of Wisconsin

There has been a marked tendency in recent years to concentrate the advertising of the better class in a few media of very wide circulation. This has led to a necessarily enormous increase in the size of the advertising section of these standard magazines in which the best and most expensive efforts of the advertiser usually appear. This increase is indicated not only by the fact that the absolute amount of advertising space per issue is steadily gaining, but by the even more significant rise in the ratio of advertising to reading space in a given issue. For example, only about 13 per cent of the total space in the Saturday Evening Post was allotted to advertising in 1900 in comparison with approximately 60 per cent in 1920.[1]

The factor of wide circulation is, of course, a requirement of primary importance in the selection of an advertising medium. Of equal importance, perhaps, is the matter of circulation among a class of readers who would be most likely to become interested in the product to be marketed. On the other hand, it is fairly obvious that any increase in the amount of advertising carried by a given medium, beyond a reasonable limit, must result in a corresponding decrease in the value of the individual advertisement, insofar as its value may rest upon the likelihood of its being seen and remembered. Indeed, it is conceivable that there is a more or less definite "saturation point" in a medium, either absolute or relative to the amount of reading matter, beyond which even a factor of such importance as very

[1] A study of certain aspects of advertising in the Saturday Evening Post, by C. J. Warden et al., Jour. Applied Psychol., March, 1926.

wide circulation might well be sacrificed in order to insure better conditions of attention and impression.

Two more or less distinct problems arise in connection with the value of the individual advertisement in sections of different size. First, there is the question of the effect of increasing the size of the section upon the chances that an average advertisement in it will be seen by the reader. Does the average reader actually turn the pages of the magazine and observe as large a percentage of advertisements when the section is large as he does when it is small? Perhaps not, and yet a definite answer must await an experimental investigation of the habits of the general public in this regard. Such a study should include in its scope both the "flat" magazine and the segregated section of the more standard type, since there is the possibility that a different reaction in this respect may be given by the same individual to the two kinds of arrangement of reading and advertising space.

The second problem is that concerning the effect of size of section (or merely the amount of advertising space in the "flat" magazine) upon the attention value and memory value of the individual advertisement when actually seen. If the reader systematically turned every page in the section and thus saw each advertisement, would the impression from a given advertisement be as vivid and lasting, on the average, when the section was large as when it was small? The average advertiser would probably answer this question in the negative from a common sense point of view, without being able to state precisely when an advertising section should be considered large or small. This factor, among others, has led the advertiser to resort increasingly to the use of skilfull layouts, novel and attractive illustrations and color combinations in order to insure a lasting impression, should the advertisement be seen at all, in spite of the competition of the balance of the space. Regardless of these attempts, however, the brute fact of the increasingly large section and its consequences remain.

The primary aim of the present investigation was to determine the immediate recall value of an advertisement in sections

of various size, and thus to obtain a quantitative measure of the value of an advertisement when it appears among few or many others.

One of the chief difficulties in making an experimental study of this problem is that of getting together a series of advertisements of approximately the same recall value to be used in the test. Advertisements differ very widely in mechanical and interest features as well as in degree of familiarity. On account of the complexity of the average advertisement it is practically impossible to equate groups, or series, in which it is used as the unit of recall, so far as the reaction value of the material is concerned. One or another group is likely to be favored by having advertisements that rank higher than the others in memory value.

In view of this fact it seemed best to carry out a preliminary experiment in which the material used should be relatively homogeneous. Instead of advertisements, small pen and ink drawings of familiar objects such as are used in intelligence tests (cow, house, bicycle, clock, etc.) were employed, all of which were very similar in size, style, and general familiarity.

Dummy magazines were made up of blank sheets of white paper of about the same size ($6\frac{9}{16}$ by $9\frac{7}{16}$ inches) and weight as the standard magazine page. A single pen and ink drawing was mounted in the center of each sheet, always on the page that would lie to the right of the subject in looking through the booklet from front to back. Dummies were made up in 25-page, 50-page, and 100-page sizes.

A chance order was followed in arranging the pen and ink drawings in the 100-page booklet. Two dummies of the 50-page size were then made up, the one being exactly the same as the first half of the larger booklet, and the other a duplicate of the second half, so that the two 50-page booklets, when taken together, contained precisely the same material, arranged in the same order, as that of the 100-page booklet. In like manner 4 dummies of the 25-page size were made, each of which was an exact duplicate of one of the four quarters of the 100-page booklet.

A total of 87 college students was used as subjects, 31 for the 100-page size, and 28 for each of the other two sizes. The subjects for the 50-page size were divided into two groups of 14 each, and one group allotted to each of the two booklets of this size, and their results combined. Likewise, for each of the four booklets of the 25-page size, a group of 7 subjects was used, and their results combined. As will be observed, the results obtained for the three-sized sections are based upon the same materials, each of the 100 objects being presented the same number of times as any other, under each size. This method should operate as an adequate control of the possible sources of error arising from differences in the makeup and familiarity of the material.

The same mode of presentation was used throughout. The subject sat at a desk on which the given section was placed. He was required to turn the pages at the rate of 2 seconds per page, observing the pen and ink drawings as he did so. The time was indicated by a metronome located out of the view of the subject. Each unit of material, regardless of the size of the section, was thus given an equal presentation time.

Immediately after the dummy had been looked through in this manner, the subject was handed a paper and pencil and asked to record, without regard to the order of presentation, the names of as many of the objects as could be remembered. The time of recall was limited to 2 minutes for the 25-page size, 4 minutes for the 50-page size, and 8 minutes for the 100-page size, and was thus proportionate to the amount of material in the section. Each subject was scored on the total number recalled, a score of one being allowed for each item.

The results are given in table 1, the absolute scores of the first column being converted to percentage scores in column three. There is a consistent decrease indicated in the recall value of an individual pen and ink drawing as the sections are doubled in size. The differences are quite large and have a high degree of reliability. The amount of loss is relatively greater as a result of doubling the 25-page size, than when the 50-page size is doubled; in fact it is more than twice as great. It would be

interesting to determine whether this apparent tendency for the rate of loss to decrease with increasing size would hold consistently through an extensive range of sections of different sizes.

The general method employed in the main experiment, in which advertisements were used instead of the pen and ink drawings of familiar objects, was the same as that described above in connection with the preliminary experiment. A few differences in detail should be noted. The booklets were composed of full page advertisements located on both sides of every

TABLE 1

Effect of size of section on the recall of familiar objects

SECTION	AVERAGE	M.D.	PER CENT RECALL	RELATIVE DIFFERENCE	P.E. DIFFERENCE	COEFFICIENT OF RELIABILITY
				per cent		
25-page	8.5	1.9	34.0			
50-page	12.8	3.2	25.6	8.4	1.7	5.0
100-page	22.3	5.2	22.3	3.28	1.2	2.8

The values in the first two columns are based upon the absolute scores for the different sized sections, while the third column shows the relative scores. The last three columns show the difference between the 25- and 50-page section, and that between the 50- and 100-page section, and the reliability of these differences.

sheet. They were not mounted and were stapled together after the manner of an ordinary magazine, so that the booklets might well have been mistaken for an advertising section taken directly from a standard magazine of the 6 by 9-inch size. Twenty-page, 40-page and 80-page sections were used. A total of 120 college men acted as subjects, 40 on the largest size, and two groups of 20 each on the two 50-page size, and 4 groups of 10 each on the four 25-page booklets.

The presentation time was 1, 2, and 4 minutes respectively for the 20-, 40-, and 80-page sections. This allowed an average of 3 seconds per advertisement, although the subjects were not

required to turn the pages mechanically at this rate. The metronome was not used since it appeared to be a distraction to many of the subjects. The recall time was limited to 2, 4, and 8 minutes respectively for the 20-, 40-, and 80-page sections. Recall of both commodity and specific brand were called for, and a score of one assigned to each, a particular advertisement being given a score of two, when both commodity and brand were recalled.

The results of this study appear in table 2 and require little comment. The third column indicates the relative recall

TABLE 2

Effect of size of section on the recall of advertisements

SECTION	AVERAGE	M.D.	PER CENT RECALL	RELATIVE DIFFERENCE	P.E. DIFFERENCE	COEFFICIENT OF RELIABILITY
				per cent		
20-page	10.9	2.8	54.6			
40-page	15.9	3.7	40.4	14.2	2.3	6.2
80-page	23.1	5.5	28.8	11.6	1.6	7.2

The absolute score is indicated in the first column and the relative score in the third column. The difference between the 20- and 40-page sections, and that between the 40- and 80-page sections appear in column 4.

value of an individual advertisement, when it appears in sections of the sizes indicated. A marked decrease in value occurs when the section is doubled in size. The amount of loss is slightly greater when the 20-page section is doubled than when the 40-page section is doubled, although the difference may have no genuine significance. The results here obtained agree in general with those of Strong as reported by Starch.[2]

In making an application of the results to the actual advertising situation, due allowance must be made for certain points of difference between the usual manner of observing advertise-

[2] Starch, Principles of Advertising, p. 796ff.

ments in a magazine section and the experimental procedure here followed. In order to secure equal presentation time per unit of material, it seems quite necessary to adopt some such method as that here used, involving a systematic turning of the pages at a fairly constant rate, although this is probably not the most usual mode of looking through an advertising section in ordinary life. In the latter instance, the time per advertisement, even when the pages are turned in order, will vary according to the interest of the reader. Then, there is the factor of the attitude of the subject which is likely to be more or less artificially determined in any laboratory situation. A definite task, not of the subject's choosing, has been set for him which requires observation according to specific rules, to be followed by a report of his observations. These artificial conditions are obviously necessary in order to control procedure and isolate the factor of size of section, and thus determine its effect upon the recall value of an average advertisement. This factor doubtless operates in much the same manner in practical life situations as we have found it to do in the laboratory.

Inasmuch as one-half of the 80 advertisements used in the above experiment were colored while the other half were black and white, a comparison of the two groups in recall value became possible. No attempt had been made, in selecting the material for the two groups, to equalize them in respect to color combination, mechanical or interest features, or degree of familiarity, and hence no genuine isolation of the factor of color was possible. An analysis of the results on this basis, in which the scores of the two groups from the sections of all sizes were pooled, gave an average recall score for the black and white group of 13.6 compared to 14.4 for the 40 colored advertisements. Since the probable error of the difference is 18.7 or only 0.042 times the difference (0.775) no significance can be attached to these relative scores.

A further analysis of the results was possible on the basis of the size of the illustration appearing in these 80 full-page advertisements. Of the 53 advertisements having an illustration less than a half page in size, only 20 (38 per cent) were

recalled 14 times or more in the pooled scores for booklets of all sizes, while of the remaining 27, with illustrations one-half page or more in size, 16 (60 per cent) were recalled 14 times or more. The average recall score for the former group was 12.1 while that for the latter was 18.6. A correlation, by the fourfold method, between the size of the illustration, as thus divided, and effective recall gave a coefficient of $+0.33 \pm 0.123$ which would seem to indicate a tendency for a large illustration to be, on the whole, advantageous to efficient recall. However, the advertisements were wholly unselected, so far as features other than size of illustration are concerned, and the result here indicated may be due, in part at least, to these uncontrolled factors.

TABLE 3

The factor of serial position in the recall of advertisements

	SECTION									
	I	II	III	IV	V	VI	VII	VIII	IX	X
20-page	28	26	21	22	20	22	20	22	19	22
40-page	13	17	14	20	14	15	17	17	14	16
80-page	11	11	14	11	14	10	9	13	12	12

The values under Roman numeral I indicate the average recall score for the first two pages of the 20-page section, the first four pages of the 40-page section, and the first eight pages of the 80-page section, in the order named. The scores for each successive tenth of each of the sections appear under the corresponding Roman numeral.

A final analysis of the data was made to determine whether or not serial position in the section was a factor in determining the efficiency of recall, under the conditions here obtaining. In the main experiment with the advertisements, the page order of each section had been systematically shifted from subject to subject, so that every advertisement appeared the same number of times as any other in each position, for the booklets of different size. The factor of serial position was thus effectively isolated.

The value of serial position in the three sized sections is shown in table 3, in which contiguous positions have been

grouped into a series of 10 successive positions in each case. Thus, the value under the Roman numeral I in the table is an average of the recall scores of the first two pages of the 20-page book; the first four pages of the 40-page book, and the first eight pages of the 80-page size, in the order named. This method of grouping was adopted as a matter of convenience in comparing the data of the three sizes of section.

The results indicate that the first 4 pages of the 20-page size are positions of especial advantage in recall which suggests the possible operation of the law of primacy in this, as in other types of serially presented material. No evidence of the operation of the corresponding law of recency is indicated, the value for the last two pages (under Roman numeral X) being no higher than for pages 11 and 12 (under Roman numeral VI) in the middle of the section. In the 40-page, and 80-page sizes, there is a considerable amount of variation in the scores for the different positions as grouped, but no tendency is apparent so far as position in the section taken as a whole is concerned. Primacy and recency are evidently of little or no importance under the conditions here employed. It is not unlikely that the unusual complexity of the units, as compared with the usual laboratory memory materials, together with the extreme length of the series here employed, may have counterbalanced any advantage arising from the operation of primacy and recency.

In any case these results suggest that the factors of primacy and recency, as these terms are commonly used in studies of memory, can hardly be responsible for the superior value commonly supposed to attach to the first and last few pages of the section, over the intermediate positions. The accepted superiority of these preferred positions more probably depends upon the fact that they are much more likely to be observed, in the ordinary course of reading the magazine, than are the more centrally located pages of the advertising section.

AN EXPERIMENT IN TESTING ENGINE LATHE APTITUDE[1]

CLYDE W. GLEASON
University of Wisconsin

This paper is a continuation of the work of Everett F. Patten, which appeared in this JOURNAL in 1923 (vol. vii, pp. 16–29). Since the purpose of the investigation is the same, and the methods the same in certain important respects, it is not desirable to describe the experiment in its entirety. If the reader is interested he may refer to the former article. The chief differences, however, in procedure and results as well as several innovations in the process, are here explained and contrasted with Patten's work.

The purpose of the investigation was to assemble a battery of tests which would be useful in the prediction of the aptitude of a beginner for learning the engine lathe. The subjects were Freshmen in the Engineering School of the University of Wisconsin, and the tests were made while the subjects still had the minimum of experience in lathe work. The criterion adopted was the quality of the work turned out in the course of the semester.

The five pieces of lathe-work turned out (four of cast iron and one of steel) were, as in the former investigation, carefully

[1] The writer begs to acknowledge the fact that the general plan of the investigation was worked out by Dr. C. L. Hull of the Psychology Department, University of Wisconsin, and that his general supervision throughout was invaluable. The writer thanks Mr. Brecky for his aid in administering the tests and especially for his careful measurement of the criteria. The several student assistants deserve credit for their care in administering the tests. He is obviously greatly indebted to Mr. Patten for the free use of the results of his work, and to Mr. Van Tassel for certain of his results, the use of which is explained later.

measured with micrometer calipers. The deviations from the blue-print specifications in each of the twenty-seven measurements taken from each student's work, were recorded, each measure subjected to check distribution. The constant median set for each of the measures was 85 per cent. The scores for each subject in each measurement were computed on a percentage basis (Patten, op. cit.). At this point the criterion was improved over Patten's work, we believe, by assigning weights of varying magnitude from one to eight (the total of weights being one hundred), the comparative weight for each measure being determined by the relative difficulty of the operation measured. The measurements were taken and the weighting estimates were made by Mr. Brecky of the Shop Practice Department. The criterion thus carefully determined was very largely objective, and, we believe, reliable.

Although Patten, in his work, obtained a multiple correlation coefficient of 0.55, it was believed that a substitution of new tests for some of those which had yielded low coefficients might appreciably raise the prognostic value of the team. Table 1 shows the tests used by Patten and those used by the writer, arranged in parallel, together with yield of zero order coefficients.

The measures taken through the use of the Coördination test apparatus (Patten, op. cit.), show, as may be seen, a decline in one from 0.42 to 0.21; in another, plus 0.19 to minus 0.16; in another, 0.36 to 0.166. Then there is an increase in the fourth from 0.03 to 0.236. The present writer's figures may be considered the more reliable for three reasons: (1) they are based upon the examination of seventy-seven subjects, whereas there were only sixty in the other investigation; (2) the "Coördination distance" measure yielded for Patten a "plus" correlation which would, as he remarks, indicate that the more deviating from the path in the test, the better the worker, whereas one would expect the opposite to be true. The present writer's minus 0.16, then, is the more likely, though not high enough to be used in the final team; (3) the low "diagonal deviation" yield, in Patten's work was not as high as he expected, and his

TABLE 1

TABLE 1

PATTEN		GLEASON	
Test	r	Test	r
Coördination test, time	−0.42	The same	−0.21
Coördination test, distance	+0.19	The same	−0.16
Coördination test, diagonal time	−0.36	The same	−0.165
Coördination test diagonal deviation	−0.03	The same	−0.236
Number checking, rights	+0.055	(Omitted)	
Number checking, wrongs	−0.035	(Omitted)	
Lock test, time	−0.20	The same	−0.05
Box test, time	−0.27	The same	−0.16
Box test, errors	−0.11	(Omitted)	
Cylinder gauging, time	−0.13	(Omitted)	
Cylinder gauging, deviations	−0.20	(Omitted)	
Match square, time	+0.16	(Omitted)	
Circle centering, time	−0.20	(Omitted)	
Circle centering, grades	−0.27	(Omitted)	
Woodworking, time	−0.119	(Omitted)	
Woodworking, deviations in dimensions	−0.095	(Omitted)	
Woodworking, deviations in angles	+0.08	(Omitted)	
		Tests added:	
		Coördination, diagonal rate	−0.07
		Cube test, time	−0.04
		Head measurements, front, combined	+0.08
		Head measurements, back, combined	+0.127
		Technical Information, grades	+0.433

prediction that a retest would raise the coefficient was realized.

The "lock assembly" test has suffered markedly each time it has been given. In a preliminary investigation by Nelson, along the same lines but with fewer subjects, the test yielded a coefficient of −0.33. Then follows Patten with −0.20, and the writer with −0.05. The test is largely one of ingenuity, and one would scarcely expect a high correlation with lathe work, which seems to require but little of that ability.

The "box directions" test (described by Patten) also suffers a drop in its prognostic value, according to results herein given—minus 0.27 down to minus 0.16. Its low correlation with the other measures saved it for final use, however, and there can be little doubt but that the test taps to a certain extent the ability to follow blue-print directions, which is so necessary in lathe-work.

The "cube" test was designed to tap what was considered a necessary constituent of lathe aptitude, namely a sense of the relationship of the parts of an object to the assembled whole, together with the ability to apply method in assembling the parts. So a group of twenty-seven small wooden cubes were placed on a glass plate before the subject, arranged in little piles, one pile consisting of those painted on three faces, another of those painted on two faces, another, on one face, and a last single unpainted cube. All of them, properly assembled would form a large cube painted on all six surfaces, with no painted surfaces in the interior. The time measure was taken, but the correlation with the criterion was very small, and the positive correlation with the lock assembly test was rather high, which latter might be expected.

A number of head measurements were secured. The instrument employed was a special cephalic radiometer designed by Dr. Hull. The instrument was so constructed that, from the mid point of an imaginary line drawn from ear to ear through the head as axis, the exact radial distances to various points on the skull may be readily measured. Among these, for example, were distances to the point of the chin, the base of the nose, and various points on the cranium. On inspection of the rec-

ords it was seen that two groups of measures, the front and back cranial, showed the most significant variations. Accordingly two front measures were combined, and two back measures combined. These two groups were correlated separately with the criterion. The coefficients of plus 0.08 and plus 0.12 are interesting and suggestive but not very valuable because of their small size. Judging from them, however, it would seem that the greater the cranial "bulge" in those directions, the greater the lathe aptitude.

The Thurstone "Technical Information" test[2] yielded by far the best results. It was believed that there should be a relationship between a young man's interest and knowledge on the one hand, and his skill in the production of the objects of his interest and knowledge, or of similar objects. It may be argued that manipulative skill leads to an interest in the things used or made, and this interest to knowledge of the facts about them; or perhaps the interest and knowledge stimulate workmanship. Possibly both are true. The test consists of one hundred statements, each relating to a technical subject on which there is more or less popular knowledge. In each statement there are four alternative words of similar nature only one of which, when used, would make the statement true. A few of the possibilities are quite obviously absurd, but most of them require some actual knowledge of the subject to supplement the common sense of the person tested. The subjects were requested to underline the most probable word of the four.

The percentage scores obtained, based on the entire one hundred answers correlated with lathe ability plus 0.39. Inspection of the sheets showed, however, that there was a considerable number of "dead" statements, in that they were too easy or too hard. To eliminate these and other non-significant statements, the writer made use of an entirely different set of data.[3] From this set of data twenty-five of the one hundred

[2] Published by the World Book Company.

[3] The writer is indebted to Mr. Richard Van Tassel of the Psychology Department of the University of Wisconsin, for his use of the Technical Information test scores of seventy-seven subjects, whose criterion grades

statements of the Thurstone test were selected as being significant, after the following fashion: The new (Van Tassel) subjects were arranged in the order of their scoring on the criterion. Their errors on each statement of the Thurston test were then recorded in a check diagram. Thus, the students were arranged in order of criterion rank along the ordinate, and the statements along the abscissa. Then the ranked subjects were divided into three sections, a high, a middle, and a low third. The error totals were then computed for each of the one hundred statements for each of the three sections of students. All statements were then eliminated which did not show a small number of errors in the high section, more in the middle, and most in the lower. This was done on the assumption that no statement could constitute a significant test of comparative ability unless, when used, the results would show the inferiors making more errors than the superiors.

TABLE 2

Van Tassel (25 statements selected): Nos.	11, 15,17,18,19,24,
Gleason (30 statements selected): Nos.	6,10, 12,15, 18,19,24,
Van Tassel:	28, 34,35,37, 43,44,46, 50, 51, 62,
Gleason:	26, 32,34, 37,38,40, 44,46,48,50,51,53,56,58,61,62,
Van Tassel:	63,68, 71,74,83,84,85, 88,100
Gleason:	68,69, 74,83,84, 86, 100

The writer's subjects were then regraded on the basis of their performance in the remaining twenty-five statements. Correlation of these new scores with the criterion yielded plus 0.433, the coefficient adopted. A higher coefficient, plus 0.474 was obtained previous to this, by using the writer's own data as the basis of selection, but such procedure unjustifiably makes use of favorable chance factors in the tester's data.

and test scores average practically the same as the writer's subjects. The data are comparable in every respect but obtained from an entirely different group of subjects.

Since the Thurstone test is printed and available for general use it may be worth while to list in parallel rows, the numbers of the statements selected upon the basis, respectively, of Van Tassel's and the writer's subjects (see table 2).

Of the twenty-five statements finally used, sixteen are common to the two groups, and those which are not, are practically all on the borderline of usability. This fact shows that there are distinct possibilities in the method, and argues for its use not alone in the administration, but the construction of such tests as well. It is understood, of course, that the statements selected as significant for this particular criterion are not necessarily so for another type of criterion.

The general conditions under which the tests were given were excellent. Each test was administered by a psychology major who was carefully trained for the task and assisted, where necessary, by another who kept the records. The Technical Information test was given to small groups of the subjects in a separate room, and the individual subjects were advanced, on completion of that test, through the rest of the series of tests, and always in the same order. The attitude of the subjects was excellent, and we had the hearty coöperation of the Engineering instructors.

The means and standard deviations of the variables (tests and criterion) finally selected are shown in table 3.

The zero order and inter-correlations are given in table 4.[4]

The P.E. of each zero order correlation is less than 0.075.

Here, as in Patten's work, the intercorrelations are low—half of them actually negative. And in this investigation, as in Patten's, more than the usual emphasis was placed upon this fact in working up the data.

The next step was to obtain the correlation of each test with the criterion, with the influence of the other tests removed, achieved through partial correlation. (Method outlined by

[4] All possible intercorrelations of the original data were computed, 86 in all. This work was greatly facilitated both in speed and accuracy by the use of the Hull automatic correlation computing machine, which was put to its first practical test in research work in the course of this investigation. The work was done in the late summer of 1924.

Yule, Introduction to the Theory of Statistics, p. 239; and Rosanow, Analysis of Mental Function p. 39–43.)

TABLE 3

VARIABLE	MEAN	STANDARD DEVIATION	UNIT OF MEASUREMENT
I. Criterion	82.258	5.703	Per cent on work pieces
II. Coördination test, period of time	1.638	0.2735	Minutes
III. Coördination test, diagonal deviation	2.812	1.706	Area in square inches
IV. Box directions, time	4.310	2.278	Minutes
V. Technical Information	14.623	4.531	Number correct, basis 25 items

TABLE 4

		VARIABLES			
		I	II	III	IV
Variables	II	−0.21			
	III	−0.23	−0.15		
	IV	−0.16	+0.163	+0.057	
	V	+0.433	−0.052	−0.275	+0.118

TABLE 5

$r_{12 \cdot 345}$ ..... −0.202
$r_{13 \cdot 245}$ ..... −0.16
$r_{14 \cdot 235}$ ..... −0.19
$r_{15 \cdot 234}$ ..... +0.413

The highest order of partials is shown in table 5.

A comparison of the above with the zero order coefficients shows a remarkably small shrinkage, which indicates how completely each test taps a separate constituent of lathe aptitude.

The coefficient of total correlation (R) was then derived and checked (by Rosanow's methods, op. cit., p. 40). The coefficients arrived at via two routes, i.e., using two sets of partials, were 0.527 and 0.524, the slight discrepancy undoubtedly due to dropping decimals. These coefficients serve as a check upon

the accuracy of the partialling and give us a useful means of expressing the total prognostic value of the tests. This is seen to be slightly (three points) lower than was obtained by Patten.

As yet the assemblage of tests cannot be considered a battery since the varying sizes of the zero order coefficients certainly points to the necessity of giving the tests widely different weightings. To this end the formulae given by Yule (op. cit., pp. 237–244) were employed, which, when worked out in detail, gave the following regression equation or forecasting formula:

$$X_1 = 84.196 - 3.782\,X_2 - 0.4958\,X_3 - 0.424\,X_4 + 0.5114\,X_5$$

In making a prediction, X_1 represents the unknown probable grade in the lathe work, of any given student, or, if you like, the probable aptitude of an applicant for a position of that nature. X_2, X_3, X_4, and X_5 represent, respectively, the raw scores of the individual in the tests of those numbers. Substitute these raw scores for the X's and solve for X_1.

As a test of the accuracy of the computation throughout, the substitutions were made by the writer for each of the seventy-seven subjects of the experiment and the estimated criterion scores thus secured were correlated with the original criterion. If there were no errors, and no shrinkage due to dropping of the decimals, the check coefficient should equal that for total correlation (here, 0.527). The check coefficient found was approximately 0.50, which was deemed sufficiently close to guarantee the general accuracy of the arithmetic.

The fact that of the seventy-seven predictions made, thirty-nine fall within three points of the actual scores, while ten are identical, serves as a rough indication of the value of the tests. Their forecasting efficiency may also be expressed in terms of percentage, the figure being determined by the use of the formula:[5]

$$E = 1 - \sqrt{1 - r^2}$$

An R of 0.50 yields a forecasting efficiency of 13.4 per cent, which compares fairly well with most tests of special aptitude.

[5] For details and explanation of this formula see article by Drs. Hull and Limp in Journal of Educational Psychology, February, 1924.

THE MEASUREMENT OF PERSISTENCE

JOHN J. B. MORGAN
Northwestern University
AND
HAZEL LUCILLE HULL
State University of Iowa

STATEMENT OF THE PROBLEM

It is recognized that the possession of a trait which is commonly called persistence may enable a person who may have deficient intellectual ability to succeed whereas one who has ability and lacks persistence may fail. The trait which is represented by this term is recognized but as yet we have not been able to give it a specific evaluation. The experiment to be reported in this paper is an attempt to formulate a clear-cut concept of persistence and to develop a means of measuring it in its varying degrees.

Persistence is a mental attitude which arises as the result of a definite environmental situation. The stimulus to persistence is some situation that would tend to hinder the free unfolding of a course either begun or planned by an individual and the attitude of persistence indicates a reaction of pugnacity toward this thwarting situation. This attitude is probably a conditioned response.[1] It might be defined as that attitude which drives a person, once he has undertaken a task, to complete it to his satisfaction.

We may state this in another way. When thwarting decreases the neuro-muscular tension of an individual we call that individual non-persistent. When thwarting increases the neuro-muscular tension of a person we say he is persistent.

[1] For an explanation of how such an attitude might be conditioned, see: Morgan, John J. B., The Psychology of the Unadjusted School Child, Macmillan, 1924, p. 66.

The measuring of this trait involves the creation of an experimental situation which must have certain specific factors which grow out of an analysis of the trait and its development.

1. The subject must be given some feeling of familiarity with the task involved. One does not persist in a totally new task for the simple reason that he has no means of evaluating the possible magnitude of the difficulty.

2. He must have experienced some success in the task. Persistence is an outgrowth of the habit of success. Hence, it is absolutely essential that the individual experience success in the type of work involved in the test in order to measure how strenuously he reacts against thwarting in that same sort of work.

3. Training in success should involve the overcoming of increasingly greater difficulties.

4. The final task must be difficult enough to thwart the most persistent person.

5. There must be some scale by means of which the varying reactions may be evaluated. Most of our test measures are in terms of degrees of achievement or the time taken to attain the same degree of achievement. Since persistence is only indicated by the subject's reaction when achievement is withheld it is obvious that the first cannot be used in a test of this sort. In a previous experiment[2] with the same apparatus, time was used as a measure, but seemed inadequate.

DESCRIPTION OF THE TEST

The apparatus used in the test was a maze constructed according to the plan indicated in the diagram. The paths were grooves cut into a maple board. The places marked A, B and C on the maze were holes so that when the subject, tracing through the maze with a stylus, came to one of these holes his stylus would drop into the hole. During the test the maze is covered with a cardboard through which the stylus is inserted. Enough of the maze is visible around the stylus to indicate the nature of the turning places but not enough to enable him to see from

[2] Loc. cit., p. 68.

one alley to any adjoining alley. He knows when he has solved any particular problem by the fact that his stylus drops into a hole.

An outstanding feature of this maze is that it can be arranged to furnish different problems of increasing difficulty. This is done by inserting blocks in different alleys, thus cutting off parts of the maze. The problems used were as follows:

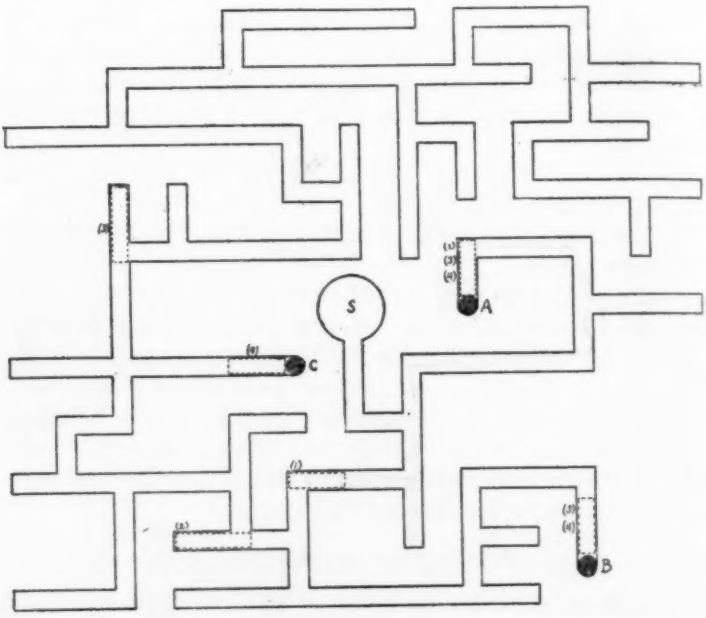


FIG. 1

1. A block as indicated by the dotted lines is inserted at point marked (*1*) on the diagram. This cuts off the greater part of the maze and the solution of this one is very easy. Hole *A* is open.

2. Blocks are inserted at the two places marked (*2*). Somewhat less of the maze is closed. Hole *A* is closed and a solution means finding *B*. This is then a new problem with a different and more difficult solution than the first problem.

3. Blocks are inserted at the three places marked (*3*). The holes *A* and *B* are both cut off and a solution involves more tracing and the location of a new hole, *C*.

4. Blocks are inserted at the three places marked (*4*). These three blocks cut off all three holes, *A*, *B*, and *C*. The solution of this maze is impossible.

PROCEDURE

The subject is seated in a comfortable position with the board, already prepared for the first test, covered with the cardboard, and with the stylus inserted at the round alley (*S*) in the middle of the board, on a table in front of him. The experimenter then says: "You see these little alleys that the stylus can travel through? You take this stylus and trace through until it drops into a hole, see, this way." The experimenter traces to the first turn and then says, "This first test is very easy, but, you see, here you can go two ways. You must choose which way to go. If the stylus does not drop into a hole, come back and try the other way. When you find the hole, tell me and I will open up more of the board for you."

Tests two, three and four are given in succession. When the experimenter opens up the board for the fourth test he should say, "Now you have practically the whole board to work on." While working the subject may say, "You must have put these pegs in wrong." If he insists the experimenter may make an examination, but it is best for him to say nothing. The subject may ask directly, "Do you know there is a hole here?" In answer simply smile. Let him interpret as he pleases.

Usually the smile is sufficient but if the subject should continually ask questions in an effort to get a pointer the experimenter may say, "I am not supposed to talk while you are taking the test." This is merely a statement of the attitude that has been preserved throughout the test, and serves to impress the subject with the worth of the test.

In giving the test the experimenter watches the manner in which the subject works and he can readily place him in one of the following nine groups:

1. Careless-anxious to quit a task.
2. Excuse hunter—always has an excuse to get out of work—feels badly, eyes hurt, time valuable, etc.
3. Fiddling plodder—keeps working because he has not enough initiative to try harder or quit, follows line of least resistance.
4. Intermittent worker—goes by spurts, working hard and then has periods of fiddling.
5. Works hard, but has little insight—(really persistent, but not intelligent).
6. Persistent worker with some method or definite attempt to accomplish aim.
7. Persistent worker with some attempt at insight but easily convinced. Probably tries out two or three methods.
8. Tenacious, obstinate type—more determined to succeed because of obstacles.
9. Analytical type—intelligently persistent to the extent that he fully analyzes the problem. Presents data well worked out to prove his conclusions.

This scale is qualitative but the distinctions between the grades are clear cut and the working on the maze brings out the different types in a very pronounced manner. In another study we are planning to work out the degree to which different observers agree in their application of the scale.

If a subject has worked a long time, the experimenter may say, "Do not work longer than you wish. If you can't find the hole, and do not care to try longer, you do not need to." A student of grade 8 type will usually work an hour or more, depending on the amount of time you have asked him to give. He will usually ask permission to work at another time. He is not yet satisfied with his work. After thinking it over, however, he does not return. One of grade 9 will return until he has mastered the test, by proving that it cannot be solved.

THE EXPERIMENT

Three separate judges were given the task of rating one hundred persons according to the scale given above. Their ratings were based on their personal observations of each of the hundred.

Three senior students were chosen to act as judges. They were superior students who had time for work outside of their

regular classes and were in their second year of teacher training, having studied psychology and pedagogy, in which courses considerable attention had been given to analyzing character traits. The nature of the work was carefully explained to them. They understood that an attempt was being made to see whether the maze test would measure the persistence of an individual, and that this would be decided by correlating the test ratings with their ratings.

Working absolutely independently each of them took the list of 100 students and carefully rated them by placing them in one of the nine groups specified on the rating sheet described above. After they had completed their ratings the twenty best and the twenty poorest individuals as determined by the average ratings of the three judges were given the maze test.

RESULTS

1. The rating sheet was an important factor in the ratings made by the three judges. A preliminary attempt to get judges to rate persistence without the help of the sheet in the final form presented in this paper gave a correlation between the judges of 0.304 ± 0.094. The correlations between the ratings of the three judges reported in this paper were as follows:

Judge 1 with judge 2	0.576 ± 0.045
Judge 1 with judge 3	0.646 ± 0.039
Judge 2 with judge 3	0.584 ± 0.044

2. Since there was such a close correlation between the ratings of the three judges it is quite probable that the selection of the lowest twenty and the highest twenty according to their combined ratings gave groups that were reliably different. The average combined rating for the lower group was 3.32 with a standard deviation of 0.92. The average combined rating for the upper group was 7.47 with a standard deviation of 0.54. The difference between the two groups was 4.14 points with a probable error of the difference of 0.163. Since the difference is 25 times the probable error of the difference it is quite evident that as far as the judge's ratings go these two groups were reliably different.

3. Having these two selected groups the question that we wished to answer was: "Will our test sort these individials out in as reliable manner as our ratings did?" Our experiment gave a definite answer to this question.

The average rating of the poor group on the test was 2.9 with a standard deviation of 1.12, while the average rating of the high group on the test was 6.65 with a standard deviation of 1.24. The difference between these two averages is 3.75 while the probable error of the difference is 0.25. The difference is 15 times the probable error of the difference. There was a slight tendency for the judges to rate them higher than the test results. Otherwise it is evident that the test sorted them out with almost as great precision as the judges' ratings.

4. In order to determine more precisely the relation between the ratings of the judges and the ratings on the test correlations were computed based on the forty cases used in the experiment. Of course a correlation based on individuals on the extremes of a scale, as our groups were, will give a fictitiously high correlation but such correlations will enable us to compare the closeness of relation if we remember that the absolute value of the correlations is not significant.

	JUDGE 1	JUDGE 2	JUDGE 3
Test	0.838	0.723	0.834
Judge 1		0.867	0.832
Judge 2			0.849

The correlations are about 0.20 higher than are the correlations between the different judges as shown under 1; hence it is likely that the correlations that we would have gotten had we tested the whole group would have ranged about 0.20 below those found by using only the extreme cases for our experiment. The correlation of the test with the combined rating of the judges was 0.842.. The probable errors of the above correlations range between 0.03 and 0.05. With the exception of the correlation of the test with judge 2, the test has as close a relation to the individual judgments as the judgments have to each

other. In the complete correlations given in 1 we see that judges one and three agree with each other more nearly than either do with judge two. Two is the erratic judge and this may account in part for the fact that his correlation with the test is lower.

SUMMARY AND CONCLUSIONS

Three judges were given the task of rating 100 pupils according to their estimates of where they fell in a definite scale of persistence which was provided them. The twenty highest and twenty lowest were then rated by the same scale but in a definite standard situation designed to bring out the trait of persistence. This persistence stimulus was an alley maze with more and more difficult problems given in succession, ending in an impossible one.

It was found that one can rate a stranger on this scale as accurately as this same person can be rated by an intimate personal acquaintance on the basis of repeated observations.

Our progress in the measurement of many mental factors has been retarded by an over-emphasis on records of quantity and quality of work produced per unit of time. Persistence seems to be a trait that cannot be measured in the traditional manner. For this reason emphasis was placed on the creation of a situation that would produce an expression of the trait, even though we had to use a qualitative measure of performance.

THE CLASSIFICATION OF MENTAL DEFICIENTS (ESPECIALLY OF GIRLS) FOR A SPECIAL CLASS CENTER

NELLIE E. POWERS

Frances E. Willard School, Boston, Massachusetts

As changes have come about in the aims of education, the process of classification of school children has taken a synthetic turn. The child and his capabilities are studied. Possible futures are being explored. The closer capabilities and possible futures can be linked up the greater will be social progress and happiness.

Mental deficients is a term used here to designate the pupils found in the public schools who are so mentally and educationally below par that they do not give promise of advancing far beyond the standards of the primary grades; still they may be partially, if not wholly, self-supporting in their later lives. The term is used in a purposed attempt to do away with the use in this connection of the term "feebleminded" with its varied and stigmatic interpretations. Even the word "deficient" carries an implication of inability to function which cannot properly, and should not be, applied to a human being until after life has given the try-out. A better term, one without a trace of implied stigma is awaited.

The term *special classes* has reference here to classes for those whose intelligence quotients range from 50 to a possible 90. It is a fairly well established fact that those whose intelligence quotients fall much below 50 are ordinarily safer under the protection of the institution. Records of the State School for the Feebleminded at Waverley, Massachusetts, show that "institutional" cases reach the institution sooner if refused admission to the special class or to the graded school. An "institutional" case, if protected in special classes until he is

beyond school age, becomes at the time of discharge a serious menace to himself and a difficult problem for society. Until classification is more carefully and scientifically worked out, the special classes, if their pupils are properly classified, can care for certain pupils with intelligence quotients of between 80 and 90 much better than the grade classes. Eventually clinic classes may better meet the need of the 80 to 90 cases; then the special classes will care for the cases with IQ's of from 50 to 80. The line of demarcation at either end of the scale should be most elastic when applied to individual cases. The intelligence quotients of the 90 special class children of this Boston study show a range of from 43 to 83. Eighty-eight per cent of the quotients fall between 50 and 83.

A *center* is interpreted here as a group of classes, largely of adolescents, whose teachers and pupils are working together coöperatively for the uplift of the group and of each individual in the group.

With the present conception of the kinds of training which are of value to pupils of such a center, three classes would be the smallest number that could coöperate in this way. Authorities differ as to the most desirable number in each class. From my own experience, it would seem that a teacher in a well organized center can work as advantageously with 18 to 20 pupils, as a teacher with a less carefully selected group can work with 12 to 15. As the number of classes is increased within a center beyond 6 or 7, the intimate contact of each child with the leader of the center is lessened. This means much more to the mentally deficient than to the normal child.

The following, then, is a study of the methods and results of the grouping of adolescent girls whose intelligence quotients lie between 50 and 90, in a small, centrally located school, where pupils enjoy, share in, and profit by experiences which promise to prove valuable to them in the emergencies of later life.

On April 2, 1924, the following questionnaire was sent to Superintendents of Schools or Heads of Measurement Departments in 141 cities in the United States. (The 141 cities were selected from those listed by the United States Bureau of

Education[1] as conducting special classes for retarded or backward or mentally deficient pupils.)

1. How many Special Class Centers for the Mentally Retarded or Deficient have you? (A Center meaning three or more classes grouped and working together.)
   Number of classes in each Center?
   Sexes, and range of ages of pupils?
   Nationalities?
2. Who chooses the pupils for a Center?
3. What is the process?
4. Are mental tests given? If so, what tests? And who gives them?
5. What are the findings from the data obtained from such tests?
6. Are printed reports available?
7. What types of pupils are admitted to a Center?
8. What types are rejected?
9. On what bases are classes within a Center divided?
10. Do you know of other Centers organized since 1922? Who is in charge of them?

Replies, very carefully worked out, were returned from 114 cities. The summary of the replies follows.

Question 1. In 48 cities, 140 centers were located. Of these centers 12 reported having more than 6 classes. The median number of classes in a center was 5. There are not as many centers for girls only as for boys, or for both boys and girls. The ages of pupils parallel the compulsory school age. There are very few American children in the classes. The foreigners are such as the location of the cities implies.

Question 2. Trained supervisors select the cases in 20 cities, psychologists or trained clinicians in 19, state examiners in 5, principals or untrained teachers in 4.

Question 3. The grade teacher, principal, school nurse or social worker first notes the backwardness. Wherever clinics are available pupils are referred to them or to a psychologist.

Question 4. Thirty-six of 48 cities report the use of or reliance upon some form of Binet-Simon Individual Intelligence

[1] Schools and Classes for Feeble-Minded and Subnormal Children. Bulletin No. 59, 1923. Published in 1924 for the Department of the Interior, Bureau of Education, Washington, D. C.

Tests. Reports from clinics show that the usual clinical procedures are followed, batteries of tests being given to each case.

Question 5. Only leading cities have made deductions from data on file.

Question 6. Very little printed matter on the subject is available. Busy workers on the problem have as yet done little writing.

Questions 7 and 8. The types of cases admitted and rejected are greatly influenced by legislation. Only 8 of 48 cities report refusal of admission to cases with IQ's below 50. Four cities frankly admit that they enter all cases that come, regardless of MA's or IQ's. Cities with strong clinics report definite limitations of 50 to 80 or 85 IQ's.

Question 9. All agree that mental age and intelligence quotient, even plus academic ability, are not criteria enough for classification. There is an emotional, or social reaction that needs more study and interpretation. The late Dr. Walter E. Fernald wrote, under date of April 29, 1924: "Most of our plans . . . . are based upon psychological findings, plus the personality and character study, plus the patient's home and other environments, etc. Every individual is a new one and his needs are just as individual." Dr. Arnold L. Gesell of the Yale Psychological Clinic writes: "The factor of vocational capacity and capacity to make social adjustments are of special importance with respect to classification."

Question 10. A large number of centers are struggling along without knowledge of what others have done or are doing.

The fact that authorities differ on such subjects as Segregation of the Mentally Deficient in Special Classes, Maximum and Minimum Curriculum, Admission Requirements for Junior High School, is evidence again of progressive thought on individual differences and individual aptitudes, and emphasizes the great value to normal individuals of studies and work on the problems presented by the mentally deficient.

On October 1, 1924, a tally was made of the sources from which 1875 pupils entered Boston Special Classes. The pupils are the population of 96 classes, including four centers.

From Grade I came 27 per cent of the cases; from Grade II, 30 per cent; from Grade III, 17 per cent; from Grade IV, 10 per cent. The remaining 16 per cent were from eleven other sources.

From much observation of placings in special classes from Grades II and III, it appears that most of the cases were probably recognized in Grade I, but were left in graded school at a parent's request. (Massachusetts law endorses parental authority in such matters.)

Terman writes: "Children are sent to school not when they are mature enough to learn what is taught, but when they have lived six times three hundred and sixty-five days."[2] As long as this is true of our American public schools, Grade I must continue to contribute the major share of candidates for special classes. Chronological age is not the best criterion of readiness to begin a school career.

Scientific diagnosis of abilities should be made at the time of entrance to Grade I. This diagnosis should be followed by immediate and authoritative placing. Parents should know and appreciate the meaning of early diagnosis.

The school[3] studied is located in the geographical center of five Boston school districts and is easily reached by electric cars from all parts of the city. The over-age children of the immediate district were carefully tested in January, 1913, by the Goddard Binet-Simon Test; 3.7 per cent of these over-age children proved to be three years or more retarded. From the retards were selected the youngest, enough for two classes. Neighboring districts contributed to another class. Three more classes were made up of pupils from any and all parts of the city. Sexes were not separated.

The pupils of the six classes were then graded according to mental age, and grouped within classes according to intelligence quotient. This grouping, although not satisfactory, continued until 1915 when it seemed best for purposes of appropriate

[2] Terman, Lewis M. The possibilities and limitations of training. Journal of Educational Research, vol. X, no. 5, December, 1924.

[3] Frances E. Willard School, Boston.

handwork to separate the sexes. The boys became the nucleus of a boys' center, and girls from distant districts filled the vacancies. The total number of girls was kept at 90, the capacity of the building.

There were great differences in ability between the twelve-year-old girl with a mental age of eight, who had sat in grade classrooms up to the fourth grade, and the twelve-year-old girl who had enjoyed fewer years in grade school and several years in a special class. Differentiations for training was necessary. Today very few girls enter the center directly from a graded school. Most of the candidates have been pupils in a single special class. Any twelve-year-old girl, pupil of a Boston special class, becomes automatically a candidate for the girls' center, though there may be good reasons which will prevent her entering the center.

Some criterion other than mental age and intelligence quotient was needed for making group divisions within the center. The healthy, overgrown girl of fourteen with a mental age of eight was as much out-of-place with younger girls whose mentality was eight, and with whom she could compete in academic and hand tasks, as she had been in the graded school. At once the problem resolved itself into a perpetual study of individual differences, individual needs, and individual placing.

When a new girl enters today she is placed in as low a class as her record makes it seem possible for her to profitably work, —in a class where she *can do* the work. This gives at once the necessary success basis, and she sees the added opportunity of several promotions ahead before she must leave the highest class when she is sixteen years of age.

The following study for purposes of classification within the center was made between January 15, 1925, and February 18, 1925. A complete study should cover physical, mental and social conditions for each child. This study omits physical diagnosis, though it does not minimize its basic importance. It is hoped that diagnoses now being made at the Boston Homeopathic Hospital Posture Clinic will yield valuable data on matters physical.

A study of the birthplaces and parentage of the 90 girls yields these facts:

Of 67 white children, 59 were born in the United States, 4 in Italy, and 4 in other European countries. All of the 23 colored children were born in America.

Of 67 fathers of white children, 18 were born in the United States, 22 in Italy, 5 in Ireland, 5 in Russia, 17 in other European countries.

Of 67 mothers of white children, 14 were born in the United States, 23 in Italy, 8 in Ireland, 5 in Russia, 17 in other European countries.

With the exception of one colored father and mother, born in England, all of the colored parents were born in the United States.

Only one child is an American of the third generation. In only three cases could mixing of nationalities be traced. Of the 59 white girls born in the United States, 51 bear names that give distinct evidence of their foreign origin.

The range of chronological ages is from 8 to 16 years. The median age is 14.5 years.

The range of the number of years the girls have been actually present in the classroom is from 1 to 11.5. The median time is 6.3 years.

The ages of many of the girls were falsely entered in Kindergarten. This alone may be a fertile source of retardation. Now birth certificates are required at entrance to school.

The low record of some girls is due to poor attendance. This is often not the fault of the girl but of her foreign parent who values a little scrubbing at home more than she does the education of her child.

A striking criticism of past classifications and of inelasticity of curricula appears, when one finds among these girls those who have been in actual attendance in school 10, 11 and 11½ years, and today have only second and third grade academic ability. They all have other abilities which, no doubt, might have been better fostered had the girls been properly placed as soon as they were recognized as academic retards. Although

special classes, and especially properly equipped centers, are very expensive, states are bearing expensive social burdens because of improperly placed retards.

Eighty-four of the 90 girls live in congested quarters. Families are large. Italian girls are living in the most crowded quarters. One Italian girl is one of ten who live in four rooms. Another is one of eleven who live in four rooms.

Splendid results of social service work appear when one studies the recreations of the children. Comparatively few of the girls play on the streets. With two exceptions, the parents welcome playmates to the home, where games are played, the graphophone entertains, and where even fudge is made. Many of the girls belong to gymnasium or dancing classes, or to clubs, connected with the Settlement Houses or with Church.

The moving pictures are the greatest outside attraction.

Of 83 girls, 70 attend moving pictures once a week, 8 twice, 4 three times, and 1 four times.

Fifty-five of 71 girls go to bed regularly by nine o'clock. Italian girls are kept up late by the late frolics of the guests of older members of the family. This may explain the fact that nearly all of the tardy marks of the school are made by Italian girls.

A test, using Haggerty Intelligence Examination, Deltas 1 and 2, proved unsatisfactory because of a seeming overlapping of the tests. However, each girl whose scores were at great variance, is an active, unstable, neurotic type of girl. One is hard of hearing. The general inability of the girls to interpret the printed page makes some of these tests unsatisfactory for these particular girls.

From the tests recommended as valuable by the various clinics, in answer to the questionnaire, four group tests were chosen, none of which presupposes the ability to read. The first given was the Dearborn Group Test of Intelligence, Series I, Examination A. Maximum score is 202. Scores ranged from 22 to 176, with the median at 109.

The second test was the Pressey Mental Survey, Primer Scale. Maximum score is 100. Scores ranged from 14 to 87, with the median at 60.

The third test was the Pintner-Cunningham Primary Mental Test. Maximum score is 59. Scores ranged from 22.5 to 55 with the median at 43.5. In the distribution of these scores most of the frequencies were 1 or 2; the largest being 6. It would seem that this is another proof that individual test and individual adjustment is necessary for intelligent classification of this group of girls.

The fourth test was the Otis Group Intelligence Scale, Primary Examination. Maximum score is 85. The scores ranged from 13 to 79, with the median at 51.5.

Every child apparently put forth her very best effort on every test. All enjoyed it thoroughly. The best work was done on tests to which there was some "carry-over" from training given in the special classes. Surprisingly few failed on the time-telling test of the Dearborn Tests. Test III of the Pressey Scale, the "Form Board Test," was perfectly done as far as time permitted. A few did not finish within the time. Test 5 of the Pintner-Cunningham Test was exceptionally successful. This is the one in which similar parts of a picture are identified. Those who did not do well on this are the over-imaginative, story-telling, or seemingly untruthful girls. The Association Test of the Otis Intelligence Scale was perfectly done by 68 of the 71 tested. For most, it was the only perfect page of the test.

As a whole these group tests do not give satisfactory ratings for the higher grade cases, nor for neurotic cases. For the steady, plodding cases, the results of group tests and of Stanford Revision of the Binet-Simon correlate highly. Even the Stanford Revision of the Binet-Simon Tests is not always dependable for accurate testing of these girls, as that test involves a certain amount of reading ability.

The general conclusion from the testing of this group of girls is that mental testing throws a great deal of light on the cases and is, therefore, a most helpful tool. It is far from being the only criterion for deciding problems of classification.

Academic measurement of the girls brought out great shortages. Courtis Arithmetic Tests, Series B, Form 1, were given

to 65 girls. An observer, unfamiliar with the imperfect correlations of effort and accomplishment, would have expected perfect results from this test. Every girl apparently gave her whole heart and soul to the work. Sixty-one rated below Grade III, 3 at Grade III, and 1 at Grade IV. Eight reached Grade III in Addition only.

As the oldest group of 16 girls was known to contain several of the school's best readers, the Thorndike-McCall Scale for Sentence Interpretation was given to them. The makers of the test state that a score of 55 marks "exceptionally low reading ability." The scores ranged from 0 to 53.

Five teachers pooled their estimates of the reading grade of the 28 best readers. The estimates were based upon the ability to read aloud, with seeming intelligence, sentences chosen from standard text books for graded schools. Then these 28 girls were rated by the Thorndike-McCall Scale for Sentence Interpretation.

Eight cases of the 28 agreed with the teachers' estimates.

Fifteen cases were over-estimated by the teacher.

Five cases were under-estimated by the teacher.

Of the 15 over-estimated, 8 were over-estimated by one grade, 7 by two grades.

The 5 under-estimated were under-estimated by one grade.

The girls whose ability was over-estimated are active girls, of intelligent appearance, who have learned to automatically pronounce words correctly. The girls who were under-rated are passive, quiet girls of less attractive appearance. The single exception to the two statements above is the case of an unattractive girl of second grade ability who appears to read aloud as well as any child of Grade V. She seldom has any idea of content.

In Handwriting, 90 girls came nearest to normal scoring. In the ratings, by the Ayres' Scale, for quality, the range was from 20 to 70, with a median of 56.6—only 3.4 short of the 60 required for "excuse from further practice."

Speed in handwriting ranged from 13 to 101 words per minute, with a median of 75. Sixty is counted a fair median for speed.

In Spelling, tested by the Ayers' Scale, the scores ranged from 0 to Grade V, with a median of Grade III.

No attempt was made to test language. Teachers' ratings range from 0 to Grade 2A. Seventy-five of the 90 girls use a foreign language at home. The remaining 15 seldom hear even fair English outside of school hours.

The following fifteen questions were selected as a basis for a social rating of each of the 90 girls:

To what extent does she

1. Respond to directions without waste of time?
2. Persevere?
3. Get along with her mates?
4. Confess wrong doings and make amends?
5. Tell the truth?
6. Stand for fairness, and play the part of a good loser?
7. Refrain from copying the work of others?
8. Show respect for property and people?
9. Give help to other children?
10. Show cleanliness in habits of person, dress, and speech?
11. Provide herself with necessary materials, keeping same, together with books, desk, and work in good condition?
12. Await her turn, avoid interrupting, remember the little courtesies, "Please excuse me," "I thank you," etc.?
13. Maintain a good standing, sitting, and walking posture?
14. Exercise self-control in keeping her temper, avoiding quarrels, refraining from sulking?
15. Practice habits of thrift in the use of money, paper, pencils, etc.?

The ratings were the medians of the ratings of six teachers. 100 represented a perfect score. The scores ranged from 23 to 100, with a median of 61.6. It is of interest to note that the 100 score was made by one of the girls with a low IQ.

The social ratings obtained by this crude scale do correlate positively with intelligence quotients obtained from the use of the Stanford Revision of the Binet-Simon Individual Intelligence Tests. The coefficient of correlation (Pearson formula) is +0.326.

The girls whose social ratings are far below their IQ's are the ones who are the misfits everywhere, the ones to whom the special class has not yet been able to make successful appeal. The girls whose social ratings are higher than their IQ's are

the center leaders. This shows very plainly in the complete set of psychographs which has been made from this study.

The graduates of the school who are now doing well at their work are those whose social ratings are high. Many a pupil with low IQ and high social rating is succeeding, while her schoolmate with a comparatively high IQ and low social rating is failing. Certain it is that the factors which make for social competency and social success offer a great field for research.

The present groups of the center, though skewed from the original MA and IQ classification, are homogeneous socially. There is much class spirit and much loyalty to the school and to its teachers. The curriculum is elastic enough to permit any girl to leave her class to fill an order for a cake, to help another girl repair her coat, to make a hat, to adjust a skirt hem, to darn a stocking or sew on a button, to fill a hot water bottle for a sick mate, to dye material for a dress, to help to scrub a floor, or to do any of the numberless pieces of practical work that need to be done where 90 girls live together.

The girls can use their hands for many kinds of useful work; and, accordingly, they are given opportunity for work in Plain Sewing, Embroidery, Dress-making, Millinery, Knitting, Crocheting, Weaving on large, old-fashioned looms, Bead-work, Paper Favor- and Flower-making, Drawing and Painting, Box-making, and Simple Book-binding. Each girl is encouraged to work for skill in the line in which she works best. Sixty-three of the 88 girls scored by handwork teachers, rank "good" or above. During the past year, over three hundred dollars worth of dresses, hats, underwear, and fancy articles were made and sold at cost prices. The money was re-invested in more materials. The workmanship compared favorably with that done at a Trade School. The handwork is all done under sympathetic, trained leadership. Perhaps some of the lack of success in life of a few former pupils may be traced to the lack of the right kind of leadership out in the world. Within the last five years great gain has been made at this point, through the employment by school departments of "follow-up" teachers, who help the employer to understand the conditions,

and help the child to see the importance of staying at her work through seeming vicissitudes. It is wisest for the present to insist, however, that the girl, upon leaving school to go to work, find her own job. The very fact that the finding of employment is difficult, and that the girl is thrown entirely upon her own resources, helps her to cling more tenaciously to the work when she finds an acceptable place.

Each girl in the school has a lesson on cookery and one in housekeeping each week. The lessons last for an hour and a half, and are conducted by a trained teacher of domestic science. In this department are made the closest contacts with life's measurements. The food must be palatable and ready on time. The dishes must be found clean and in their proper places. The room must be tidy. Much of the work is known to carry over directly into the home.

Lunches are cooked and served at cost at the school each day for from 60 to 75 pupils and 6 teachers. Pupils of the oldest classes in groups of four, in turn, prepare the lunches. All classes take turns, for a week at a time, in clearing up the room after lunch. The work parallels the work done in domestic science in the graded schools with but few exceptions. Special class girls fail on tasks requiring academic ability. They do not, as a rule, copy recipes correctly. They are seldom sure of measures. They must be supervised at just the right moment in just the right way. Few girls can purchase materials, make and frost a cake, without a little help. Nevertheless the school has a reputation abroad for its "splendid cakes." The reputation finds its source in the psychology of the mentally deficient. The thorough beating of the ingredients is basic for successful cake making. Deficient girls are happy when they can use the egg-beaters, and will beat tirelessly until told to stop. The teacher's part is the giving of the "stop" signal, followed by the next simple direction, at just the right time. Surely mental deficients have a place of real worth in the world, if it be no higher in the scale than beating cakes.

Included in the study are psychographs of each girl, showing her Chronological Age, IQ, School Grade for present CA, School Grades in Arithmetic, Reading, Spelling, Writing, Hand-

work, Cooking and Housekeeping, and a Social Rating. The value of the graphs will develop with repetition of tests at regular intervals.

A study of the placing of 321 girls discharged from the school in the past twelve years shows that fewer criminal cases and fewer institutional cases appear in the classes than formerly; fewer are returned to grades because of parental unrest. More and more are the graduates employed in factory, or remain at home to work while others of the family earn the daily bread.

This confirms a statement made by the late Dr. Fernald in a lecture before the Boston Special Class Teachers on October 3, 1924: "More and more are the special classes keeping out of the institution the morons who have ability to do some useful, real work worth a wage."

In an attempt to locate 101 of these factory and home workers, 31 were found in factory earning from $7 to $18 a week. Many, especially the colored girls, are unemployed. Of the 21 found at home only 6 have really done well there. The home supervision is often lax, and suffering and disease have crept in. These home girls are the low grade girls. More institution room might relieve this situation.

This is Dr. Fernald's thought: "In order that the mentally handicapped may be an asset and not a liability, early diagnosis, appropriate training, kindly social supervision are imperative; and the organization of the requisite provisions for these three activities devolves on the state and on the community at large."[4]

Early diagnosis is already possible. Great advances are being made in making the training appropriate. Practical work on social supervision of special class pupils after leaving school is still in its infancy. The next step beyond the work of "Centers" will undoubtedly be some kind of "actual work" or "shop" schools for girls and for boys, where those with capable hands can be more practically tried out at the different lines of work which "follow-up" shows are the actual lines of real life activities. Classification is basic at each and every step.

[4] Boston Medical and Surgical Journal, vol. 192, no. 2, pp. 94–96, January 8, 1925.

AN INFORMATION TEST APPLIED TO JUVENILE DELINQUENTS[1]

EVELYN EASTMAN

Wayne County Psychopathic Clinic, Juvenile Court, Detroit, Michigan

One of the most frequent and valid objections to the Binet Scale in its several revisions is that it does not measure social efficiency, the lack of which is extremely important in mental diagnosis in detecting the potential social inefficient. Very often the recommendation, especially in cases of borderline intelligence, of all diagnostic levels the most uncertain, depends upon the degree of social fitness and estimated industrial capacity. An inquiry by means of a questionnaire covering allopsychic reactions is a rapid but effective method of supplementing the Binet Scale in this respect and in aiding the examiner in his recommendation. On the Stanford revision there is no test item which is definitely reliable in furnishing a knowledge of the subject's awareness of his relation to his environment above year IX. That orientation has a very direct bearing upon the social efficiency of the individual is a matter of common acceptance; in some cases of cursory or preliminary examination or for confirmatory evidence it is the only method employed. By the layman, its lack in an appreciable degree is regarded as almost an undisputed criterion of an unintelligent being. The use of the Ten Point Scale of Psychopathy[2] in many clinics involves the determination of orientation as an aid in disclosing the psychopathic function. To this end, desiring an instrument of sufficient range to lend interpretation to variations in score, the director of the Wayne County Psycho-

[1] Submitted and accepted in partial fulfillment of the requirements for Graduate credit in Applied Psychology.

[2] Florence Mateer, Ph.D., Journal of Delinquency, Vol. VI. No. 1, January, 1914.

pathic Clinic,[3] with the collaboration of her staff, devised an Information Test of 100 questions to be applied to individuals above eight years of age, and the study herewith presented is an analysis of the results of the first 300 such tests given to unselected patients. Such a test, intended to be applicable to all ages and to people of varying educational attainments, is necessarily of local interest in its subject matter; however, in statistical analysis and interpretation of results there are no geographical limitations for in these its interest ceases to be wholly local.

The test consists of 100 questions, requiring about ten minutes time for its administration, each item of which is assigned an arbitrary scoring value of one point. It is reproduced in original form at the conclusion of this article, the more recent changes being omitted inasmuch as they are not included in this study of the first 300 cases to whom it was given.

The chronological ages of the subjects and the corresponding scores are given in table 1 following the general plan of a twofold grouping, those whose IQ's were above 80, numbering 132 cases, and those whose IQ's were below 80, numbering 168 cases; in the former group there were 50 boys and 82 girls, in the latter, 69 boys and 99 girls. In each group there were several cases whose chronological age would place them in an adult classification, but because there were so few of them they were considered under one comprehensive category, designated as "above 16." This table shows a disparity between the median scores of the two groups which is never closer than 11 points for those whose chronological age is above 16, and varying 34 points in years 9 and 11. The general tendency is for the variance to become less as the chronological age increases and we may thus conclude that a deficiency in orientation and general information tends to lessen in proportion to the intellectual defect; that is, a child whose IQ is below 80 at nine years of age will have an information test score 34 points below that of a normal child of similar age, but when the two have attained

[3] Dr. Nellie L. Perkins.

the age of 17 there will be but 11 points difference in their scores, theoretically.

MENTAL AGE VERSUS SCORE

The mental ages of the higher group range from 7 to 17 years, the median being between 12 and 13 and the median score 87. The median score of the inferior group is 63 and the median mental age between 8 and 9, range 6 to 12.9 years. The consistent relationship between mental age and information

TABLE 1

I.Q.'s above 80			I.Q.'s below 80		
Chronological age	Score	Number of cases	Chronological age	Score	Number of cases
Above 16	94	7	Above 16	83	9
16	94	17	16	74	27
15	93	30	15	71	41
14	89	18	14	70	37
13	84	24	13	58	22
12	82	11	12	58	14
11	70	6	11	36	4
10	58	7	10	35	8
9	59	6	9	25	5
8	39	6	8	15	1
		132			168

test score is shown in tables 2 and 3 in which the median scores by year groups are progressively lower as the mental age decreases.

SCORES VERSUS DIAGNOSES

The patients had previously been classified according to The Statistical Manual prepared by The National Committee for Mental Hygiene and it was possible to compare the ratings of the different diagnostic groups. As might be expected the feeble-minded made up the largest part of the inferior group, those with IQ's below 80; while the group designated "no psychiatric condition" made up the greater part of the superior

TABLE 2

Score versus mental ages

I.Q.'s ABOVE 80		I.Q.'s BELOW 80
Mental age	Score	Score
17.0	96	
16.9	93	
16.6	92	
16.3	94	
16.0	93	
15.9	91	
15.6	85	
15.3	95	
15.0		
14.9	94	
14.6	96	
14.3	95	
14.0	91	
13.9	91	
13.6	92	
13.3	89	
13.0	88	
12.9	91	90
12.6	89	90
12.3	84	79
12.0	76	80
11.9	74	89
11.6	84	78
11.3	83	79
11.0	85	83
10.9	74	74
10.6	72	74
10.3	71	73
10.0	64	73

TABLE 2—*Continued*

I.Q.'s ABOVE 80		I.Q.'s BELOW 80
Mental age	Score	Score
9.9	65	66
9.6	62	63
9.3		62
9.0	60	58
8.9	48	75
8.9	58	57
8.3	44	55
8.0	46	62
7.9		53
7.6	28	46
7.3	48	30
7.0	26	36
6.9		22
6.6		24
6.3		22
6.0		

TABLE 3

Median scores by year groups

I.Q.'s ABOVE 80			I.Q.'s BELOW 80		
Number of cases	Mental age	Score	Number of cases	Mental age	Score
19	16	95	21	12	84
7	15	92	19	11	83
17	14	93	32	10	74
24	13	91	43	9	61
23	12	35	22	8	58
10	11	82	23	7	43
7	10	71	8	6	22
9	9	62	—		
12	8	50	168		
4	7	29			
—					
132					

group, those with IQ above 80. With but two exceptions in years 12 and 13 the constitutional inferiors rated higher than the subnormals and the subnormals higher than the feeble-minded. Likewise, with but one exception, the constitutional inferiors whose IQ's were above 80, scored higher than their corresponding group whose IQ's were below 80, although the diagnoses were the same. The psychoneurotics fell only within the superior group, and although they were numerically small, their scores were higher than those of any other diagnostic group.

TABLE 4

Boys' versus girls' scores

I.Q.'s ABOVE 80					I.Q.'s BELOW 80				
Age	Boys		Girls		Age	Boys		Girls	
	Score	Number	Score	Number		Score	Number	Score	Number
Above 16			91	7	Above 16	59	3	76	6
16	93	3	93	14	16	81	8	71	19
15	96	11	91	19	15	61	13	76	28
14	93	5	87	13	14	71	15	71	22
13	83	7	81	17	13	60	8	58	14
12	89	6	81	5	12	68	7	40	7
11	78	3	72	3	11	36	4		
10	73	4	62	3	10	40	5	33	3
9	67	5	63	1	9	25	5		
8	62	6			8	15	1		

The psychopaths, on the other hand, generally fell below the scores made by the no psychiatric condition group, although invariably higher than the constitutional inferiors. Our results, then, do not confirm the Ten Point Scale of Psychopathy, in that a high score does not indicate the psychopathic function.

SEX DIFFERENCES

Table 4 can lead to but one conclusion—that boys score higher than girls in a test of this kind; this is particularly manifest in

the better group, IQ above 80. An analysis was not made of the nature of the questions failed by boys as against that of girls, but it may be presumed that boys, being less restricted in their freedom on city streets, acquire a more extensive knowledge of their immediate environment.

SCORES VERSUS SOCIAL GROUPS

In the estimation of social status, an arbitrary value was assigned by qualifying the total home influences as poor, fair, or good. A further check was obtained by designating the education of the parents on a Scale of 10 points, obtained as follows:

0—Illiteracy
1—Literate
2—Grade school
3—Grammar school
4—High School
5—College

The information covering educational background was not accurate enough for anything more than this general rating. Analysis showed a greater tendency for the scores to increase in proportion to the higher rating of the home and educational advantages of the parent, indicating that the child of inferior intelligence from a home possessing good features and of parents educated beyond mere literacy, is apt to be better oriented and better informed than his companion in a poorer environment. To some extent this is true of the child of the superior group, similarly situated, but the general indications are less marked that he will surpass his less fortunate brother in poorer surroundings.

The only two noteworthy features of the analysis of scores by school grades (table 5) are (1) that the progression of median scores is consistently gradual from the second to the eleventh grades, the chief break occurring after the fourth grade, and (2) that the score attained by Special Room pupils is practically equivalent to a fifth grade score. It must be recalled, however,

that the Special Grade score represents older pupils who are retarded while the fifth grade score represents also a certain number at age for grade. Comparing this table with table 1 it will readily be seen that the scores for chronological ages are higher than the corresponding grade scores to be expected, the cause for which discrepancy is probably due to the fact that pedagogical retardation is as a rule to be found among these clinic cases.

TABLE 5

Score versus school grade

GRADE	MEDIAN SCORES	NUMBER OF CASES
Second	39	4
Third	51	32
Fourth	66	24
Fifth	77	32
Sixth	80	20
Seventh	83	36
Eighth	92	52
Ninth	96	44
Tenth	97	12
Eleventh	96	8
Special	75	28

LANGUAGE DIFFICULTY VERSUS SCORE

Language difficulty, among our cases, occurred only in Polish, Italian, and German speaking homes. The cases in which language difficulty was noted and recorded by the examiner are only 48 in number, which when tabulated by various years, nationalities, and degree of intelligence, do not present reliable statistical facts. However, it may be stated that in general the amount of deficiency on an Information Test does not seem proportional to the handicap which is encountered on the individual intelligence test. That is, a child more quickly acquires street jargon and information concerning his immediate environment than he does the language necessary for the more formal test material.

TABLE 6

Tests failed by group whose IQ's are above 80

1. Assumption is made that what is failed by a 16 year old will, as a rule, be failed also by those younger.
2. By failure is meant that 75 per cent or more of group at a given age fail to pass test.
3. Items are given in order of difficulty.

Year XVI.	Name something made in Battle Creek. In Grand Rapids. Name 2 law courts in Detroit. Who is Frank Cody? Location of Post Office and County Building. Location of City Hall and Main Library.
Year XV.	What is meant by registering letters?
Year XIV.	Who is James Couzens? How can we get from here to town by streetcar?
Year XIII.	What is meant by jay walking? What does W. W. J. stand for? Who can be a citizen? What is meant by insuring a package? Give one safety first rule. What does D. S. R. stand for?
Year XII.	On what continent do you live? Who is Edgar Guest? What is the capital of Michigan? What is meant by sending a letter special delivery?
Year XI.	What was your mother's maiden name? Name a residence street in Detroit. A boulevard. A business street. Name the months of the year. What is the capital of the United States? What is a bootlegger?
Year X.	In what year were you born? In what county do you live?
Year IX.	In what country and state were you born? How tall are you? What would you do if you were lost?

TABLE 7

Tests passed by group whose IQ's are below 80

1. It is assumed that what is passed by an 8 or 9 year old is also passed by those older, therefore the additions to each year are cumulative.
2. By passing a test is meant that 75 per cent or more of the group at a given age succeeds with it.
3. Items are given for various years in order of ease of accomplishment.

Years VIII and IX. Name. Age. Address. Name of city.

Year X. Names of brothers and sisters.
Are your parents living?
Who brought you to this clinic?

Year XI. Have you a bank account?
Name 2 newspapers.
What is your father's name?
Name the days of the week.

Year XII. All of the above items plus:
How long have you lived in the U. S.?
Name 2 things you must do in sending a letter.
What does your mother do?
Where can you buy gasoline and milk?
What is a janitor?
How much do you weigh?
Name the months of the year.

Year XIII. All of the above items plus:
What does your father do?
Who is Harold Lloyd?
Names of 3 street car lines.
What should you do if a fire breaks out in your home?
Names of 2 hospitals.
Name of one church.
What does U. S. stand for?

Year XIV. Where do you buy a hot water bottle and screw driver?
Name something made in Detroit.
When is your birthday?
Where were you born?

TABLE 7—*Continued*

Year XV.	Name 5 street car lines in Detroit.
	Who is Henry Ford?
	Name a business street in Detroit.
	Name 3 boat trips you can take from Detroit.
	Name 2 moving picture houses.
Year XVI.	Why are you here?
	What is a conductor and janitor?
	Give one safety first rule.
	Where can a bolt of cheese cloth be bought?
	What would you do if you were lost?
	Names of 3 Detroit stores.
	Hour of day and date.
	Name of state.
	In what year were you born.
	Who is Andy Gump?

CRITIQUE OF THE TEST

An analysis of the difficulty and ease of accomplishment of individual items is presented in tables 6 and 7. As would be expected in a test applied to individuals ranging in chronological age from 8 years to adulthood, not all questions would have the same practical value. Nevertheless, there are two extremes wherein some of the tests have no value as differentiating instruments, these to be found in those tests passed by more than 75 per cent of the 8 and 9 year olds and those failed by 75 per cent or more of the 16 year olds. On the basis of these findings, the test is being modified and reorganized to eliminate valueless questions, and perhaps to reduce its length if compatible with a range of sufficient variation to allow for interpretations.

SUMMARY AND GENERAL CONCLUSIONS

1. There is a general tendency for the degree of orientation and amount of general information which a subject possesses to increase disproportionately as he becomes older; i.e., he acquires more than his intellectual status would predict.

2. Scores decrease as the mental age decreases for given age groups.

3. For diagnostic groups, the order of arrangement in respect to scores is arranged in descending rank; psychoneurotics, no psychiatric condition, psychopaths, constitutional inferiors, subnormals, feeble-minded. Our tests do not show that a high score indicates the psychopathic function.

4. Boys, particularly in the higher group of IQ's, rate consistently higher than girls.

5. Better home environment and superior educational advantages act to give the child of inferior intelligence similar advantages in orientation and information; the same is not true of the normal or superior child who seems to rate equally well regardless of environment.

6. Gain in score is consistent with progress in school.

7. Language difficulty is less operative on a general information test than upon more formal tests requiring linguistic facility.

WAYNE COUNTY PSYCHOPATHIC CLINIC

D. P. C. No.....Name.....Date.....Examiner.....

General Information Test

1. (a) What is your name?
   (b) How old are you?
5) (c) When is your birthday?
   (d) In what year were you born?
   (e) Where were you born?

2. (a) What is your address?
   (b) In what city do you live?
   (c) In what county do you live?
8) (d) In what state do you live?
   (e) In what country do you live?
   (f) On what continent do you live?
   (g) What is the capital of Michigan?
   (h) What is the capital of the United States?

3. (a) How long have you lived in the United States?
   (b) How long have you lived in Michigan?
   (c) How long have you lived in Detroit?
   (d) What year is this?

9) (e) What month?
(f) What day of the month?
(g) What time is it?
*(h) What is meant by Standard time?
*(i) In what time belt is Detroit?

4. (a) What are the names of your brothers and sisters?
(b) What are the ages of your brothers and sisters?
(c) What was your mother's maiden name?
6) (d) What does your mother do?
(e) What is your father's name?
(f) What does your father do?

5. (a) Name five car lines in Detroit.
(b) How could you get from here to town by street car?
(c) Name three big Detroit stores.
13) (d) Where is the Main library?
(e) Where is the City hall?
(f) Where is the main Post Office?
(g) Where is the County building?

6. Who is each of the following?
Ty Cobb
Henry Ford
Edgar Guest
10) Andy Gump (Detroit Free Press comic)
Harold Lloyd
James Couzens (former Mayor, present Senator)
Frank Cody (Superintendent of Detroit Public Schools)
What is meant by
D. S. R. (Detroit Street Railway)
W. W. J. (Detroit News Radio Station)
*W. C. X. (Detroit Free Press Radio Station)
7. (a) Have you a bank account?
*(b) How would you start a bank account?
7) (c) What are the hours for banking?
*(d) Between what hours are stores generally open?
*(e) Name two things you could do at the Post Office.
*(f) Where could you go for help from the City?

8. (a) If you were lost what would you do?

---

* This question has been dropped in the last modification because it did not meet the requirements of the test.

(b) What would you do if a fire broke out in your home?
(c) What is jay walking?
7) (d) Give one safety first rule.
(e) Name a business street, a boulevard and a residence street of Detroit.

9. (a) Where could you buy a hot water bottle; screw driver; bolt of cheese cloth; gasoline; milk?
(b) What is a janitor?
What is a conductor?
What is a bootlegger?
27) What is a carpenter?
What is a traffic officer?
(c) Name one thing made in
Grand Rapids
Battle Creek
Detroit
(d) Name two newspapers published in Detroit.
(e) Name two courts in Detroit.
(f) Name four hospitals.
(g) Name three boat trips you could take from Detroit.
(h) Name two moving picture houses.
(i) Name one church building in Detroit.

10. (a) How tall are you?
(b) How much do you weigh?
(c) Are you a citizen?
(d) Who can be a citizen?
8) (e) How can one become a citizen?
(f) Who am I?
(g) Who brought you to this clinic?
(h) Why are you here?

A STUDY OF SOME COMMONLY USED PERFORMANCE TESTS

MYRTLE RAYMAKER WORTHINGTON

Institute for Juvenile Research, Chicago, Illinois.

The following study is an analysis of eleven commonly used performance tests, to determine whether these tests contribute any additional information to the general intelligence rating measured by the Stanford-Binet scale, and to determine the best substitutes to be used when the Stanford-Binet or other intelligence test cannot be applied.

In the clinical examination of children, many types of supplementary tests are ordinarily given in addition to tests of general intelligence, such as vocational tests, educational tests, volitional and emotional tests, and a large number of performance tests, few of which have stated purposes. Many of these tests are used to verify the mental rating, particularly when there is language difficulty, deafness, or other defect which might invalidate the Stanford-Binet rating.

The common assumption that many of the various performance tests used furnish information additional to that given by the Stanford scale has not been verified. Necessarily, in any individual examination, only a limited number of all the known performance tests can be used and the question of selection is important. Thus far, many of the tests are commonly considered as attention tests, alertness tests, tests of social adaptability, etc.[1]

[1] In an article by Garrett, H. E., and Lemmon, V. W.: An analysis of several well-known tests, Jour. Applied Psychol., 1924, viii, 424–439, the author presents the same need for analysis of tests that we do. The work was done with college students and the only one lapping of tests used is with the cancellation of A. He finds a correlation of 0.40 with intelligence.

The specific performance tests selected for study were the following: Healy Picture Completion's I and II, Healy Mare and Foal Picture Board, Healy Construction Puzzle A, Seguin Form Board, Knox Cube Test, Glueck-Knox Ship Test, Porteus Maze, Cancellation of A, Pintner Directions Test, and the Trabue Completion Test.

Healy (1) Picture Completion I was originally devised by Healy for "estimating the power to make connection between different portions of the mental content," (p. 189). Dr. Healy found that "at ten years the performance is as good as at thirteen years . . . . in fact, at nine years the results are not far behind" (p. 201). He states in conclusion, "We evidently have in our completion picture a test for ability primarily adapted to the child type of mind. . . . . We cannot expect a much better average performance with increasing years; indeed, good results may not be obtained from older cultivated individuals" (p. 203). Pintner and Anderson (2) standardized this test for mental ages five to sixteen inclusive, and their norms were used in this study.

Healy Picture Completion II (3) was devised by Healy as a test of apperception, considering apperception as essential to "any fair attempt to measure intelligence." This test was standardized by him for ages six to sixteen inclusive.

Healy Construction Puzzle A (4) was devised "to bring out the perception of relationships of form and also the individual's method of mental procedure for the given task, particularly for his ability to profit by the experience of repeated trials in contradistinction to the peculiar repetition of impossibilities characteristic of the subnormal and feeble-minded groups." He states that one should record the time and number of moves made, but gives no norms. This test was standardized by Pintner and Paterson (5) on 1005 school children, ages seven to sixteen, according to moves and time. Since both norms are considered as equally reliable for mental age level, norms for time only were used in this study, time score being more easily obtainable.

Healy Mare and Foal Picture Board was devised by Healy

(4) to test "perception of differences in form, powers of coördination,—the ability to learn by the experiences of trial and success." Pintner and Paterson (6) simplified this test by omitting the three geometrical figures, and standardized it for ages five to sixteen inclusive, according to speed of performance, on 667 school children.

In the use of the Seguin Form Board, a test for the perception of form relations, Sylvester's (7) standardization was adopted. Norms were available for ages five to fourteen inclusive, according to the shortest of three trials.

The Knox Cube test was originally used by Knox (8) in his scale for estimating the intelligence of non-English speaking adults, and is presumably a test of general intelligence. In this study, Pintner's modification of the test was used, norms being available for ages four to fifteen inclusive (9).

The Ship test was originally devised by Glueck as a test of general intelligence in non-English speaking adults, and was adopted by Knox (10) in his scale for mental defectives, using the test at age levels thirteen and above. The test was standardized by Pintner and Paterson (5) for ages five to twelve inclusive. The norms furnished by them were followed in this study.

Porteus (11) originally devised his Maze test as a test of social fitness, which he states, demands "ordinary prudence, foresight, and common sense methods," and the ability "to sustain attention and to succeed in a task." The test is standardized for ages three through fourteen.

The norms used for the Cancellation of A test (hplg form) were those given by Pyle (12). He states that the test measures "quickness of perception and discrimination and to some extent quickness of reaction" . . . . and adds, "as far as I have been able to determine, it has not very close relation to general intelligence."

The Pintner-Toops Revised Directions (13) test used is a modification of the Woodworth and Wells Easy and Hard Directions (14) test. These latter tests were devised by Woodworth and Wells to "attempt to test the ability to understand instruc-

tions—that is, to discover the subject's speed in apprehending them." It was assumed by Pintner and Toops that the test measures what the originators intended it to test, and they made no further analysis of it. They standardized it on 1756 cases, children and adults.

Scale A of the Trabue Completion test (15) was used in this study. Concerning this test the author says, "On the whole, however, it will be found that ability to complete these sentences successfully is very closely related to what is usually called 'language ability.' Ebbinghaus characterizes it as 'a real test of intelligence.' It has been variously classified by psychologists as a test of 'imagination,' as an 'association test,' as a 'test of memory,' and as a test of numerous other 'faculties'; as a matter of fact, the completion of incomplete sentences correlates remarkably well with almost any other measure of desirable qualities. It would perhaps have been just as descriptive to have used the classification 'general stand-out-ishness,' suggested by Professor Hollingworth." Trabue classifies it as "for educational purposes as 'language scales.'" Norms were available in this test for ages four to eighteen.

The performance tests described, together with the Stanford-Binet intelligence test, were given to 539 subjects. Because of time limitations, the entire series of performance tests were not given in each case. Since the purpose of this study is primarily the examination of individual tests, and not the individual performances, results are not affected by incompleteness in series. The subjects for this study were unselected cases examined at the Institute for Juvenile Research,[2] children referred for examination because of mental defect or behavior difficulties. The distribution thus covered a wide range of chronological and mental ages. This distribution is given in table 1.

The distribution curve made on the intelligence quotients ranged from below 0.50 to above 1.39. The curve was skewed on the lower ends, since approximately twenty-five per cent of

[2] The Institute for Juvenile Research, Dept. of Public Welfare, State of Ill., Chicago, Ill. Herman M. Adler, Director.

the cases fell into each of the following groups: 89–109; 79–89; 69–79, i.e., the intelligence of this group falls just below a normal, unselected group.

The method of study of individual tests used was to find the correlation of their mental ages with the Stanford-Binet mental age, assuming that the Stanford-Binet is a measure of general intelligence. All correlations were further checked by repeating the tests on approximately fifty new subjects. If a test

TABLE 1

MENTAL AGE	CHRONOLOGICAL AGE														
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
3	1		1	2		3	1		1						9
4			7	7	4	2		1	1	1					23
5			2	13	4	7	5	2	1		3	2	1	2	42
6			2	4	5	9	11	3	3	1	1	1		4	44
7			1		7	13	11	8	7	2	2	2		8	61
8					3	6	13	19	12	8	4	3	6	12	86
9						3	8	14	9	16	12	15	9	9	95
10					1	1	2	8	7	6	13	13	4	9	64
11									3	7	2	12	8	10	42
12									3	4	4	7	3	4	25
13									1	2	2	7	8	5	25
14								1		1		5	3	1	11
15									1		1		2	2	6
16								1				1		1	3
17														2	2
18														1	1
	1	0	13	26	24	44	51	57	49	48	44	68	44	70	539

proves to have a low correlation with the Stanford-Binet, it is evident that the test does not measure general intelligence; if the correlation is high, it may be concluded that the test measures at least some factors of general intelligence, and is, therefore, valuable as a substitute for the Stanford-Binet, where the latter cannot be given. Table 2 gives the correlations obtained on the first series of cases, as well as the correlations found by Johnson and Schriefer (16) on some of the same tests.

The Johnson and Schriefer correlations were made between the chronological ages of the children and the mental ages on the performance tests.[3]

Since these correlations may have been affected by the fact that the unselected cases examined were really selected as far as intelligence was concerned, i.e., the majority of the cases had a mental age of thirteen years or below, a second series of fifty cases were tested.[4] The second correlations verified the findings of the first tests, i.e., although correlations were, on the

TABLE 2

TEST	NUMBER OF CASES	CORRELATION	P.E.	JOHNSON AND SCHRIEFER
1. Pintner-Toops Directions	177	0.79	0.001	
2. Seguin Form Board	346	0.76	0.022	0.65
3. Porteus Maze	118	0.75	0.03	
4. Trabue Completion	123	0.73	0.029	
5. Healy Picture Completion I	118	0.60	0.039	0.55
6. Healy Construction Puzzle A	109	0.59	0.041	
7. (Healy Mare and Foal)	124	0.55	0.043	0.55
8. Glueck Ship	119	0.46	0.048	0.53
9. Healy Picture Completion II	143	0.41	0.047	
10. Cancellation of A	131	0.46	0.046	
11. Knox Cube Imitation	122	0.56	0.042	0.51

whole, lower, due to the smaller number of cases, yet the respective places were the same and the type of distribution remains the same.

[3] A study by F. Gaw: A study of Performance Tests, Brit. Jour. Psychol., April, 1925, presents results with fourteen tests. The work was done with English Elementary School and Canal Boat Children. One of the conclusions is that performance tests are less influenced by environment and lack of schooling than is the Binet.

[4] The following correlation on fifty cases were found: Pintner-Toops Directions, $r = 0.83 \pm 0.03$; Seguin Form Board, $r = 0.55 \pm 0.01$; Porteus Maze $r = 0.59 \pm 0.03$; Trabue Completion, $r = 0.87 \pm 0.02$; Healy Picture Completion I, $r = 0.57 \pm 0.06$; Healy Construction Puzzle A, $r = 0.42 \pm 0.07$; Mare and Foal, $r = 0.42 \pm 0.07$; Glueck Ship, $r = 0.49 \pm 0.07$; Healy Completion II, $r = 0.42 \pm 0.08$; Cancellation of A, $5 = 0.37 \pm 0.09$; Knox Cube Imitation, $r = 0.48 \pm 0.07$.

The first four tests in table 2 shows a "high" correlation with the Stanford scale. These are the tests which showed a regular distribution, with no disproportionate number of cases making unusually high scores on the test. These results indicate that these tests, therefore, may be used as tests for determining the mental age level, and are not tests which measure any special abilities, other than those included in general intelligence.

The Pintner-Directions test has been described as a test of the ability to understand instructions from the high correlation with Stanford, we may assume in this case also that the ability to understand instructions is a factor of general intelligence.

The high correlation of the Seguin Form Board with Stanford indicates that this test is a valuable one to use as a supplementary non-language performance test.

It has been said that the Porteus maze is a measure of social adaptability, considering this trait independently from intelligence, but the high correlation of 0.75 with Stanford indicates that if the test is really a measure of social adaptability, then this trait is a function of general intelligence.

The Trabue Completion scale is purely a language test, so that one would expect a fairly high correlation with Stanford. The correlation of 0.73 is sufficiently high to indicate that this test may safely be used to give a rough estimate of the intelligence level, when only a brief time interval is available, or when a largé group is to be examined.

The fact that these four tests correlate highly with Stanford is not an indication that these tests alone are equivalent to the Stanford scale. In the first place, at best these tests give mental ages only within half year groups, and in the second place, they do not give the qualitative differences shown by means of the Stanford Examination.

The next four tests listed, with correlations of 0.60 or below, all showed irregularities of distribution, that is, many subjects from six to twelve years made scores as high as sixteen years on certain of the tests. An illustration is given in table 3, which shows the distribution table for the Healy Construction Puzzle

A. When the distribution is characteristic of that shown in table 3, it would indicate either that the test is poorly standardized or that it measures some special ability other than general intelligence. In the latter case it would seem that the test should not be used to indicate general mental age level, comparable to the Stanford, but should be used only for the purposes of qualitative analysis. This seems to be the case with the Healy Picture Completion I, Healy Construction Puzzle A, Healy Mare and Foal, and the Ship Test. Healy, when he devised his tests, evidently meant them to be used for purposes

TABLE 3

HEALY A MENTAL AGE	MENTAL AGE, STANFORD-BINET													TOTAL
	4	5	6	7	8	9	10	11	12	13	14	15	16	
16		2			3	1	5	1	4	2	2	1	1	22
15			1		1	2	3	2	2	2	2			15
14						1	2	1						4
13					2		1							3
12			1		2		1							4
11			1	4	1	1	1							8
10	1	1		5	2	3	3	1						16
9		2	2	2	1		1	1						9
8				1	1	5								7
7	1	4	2	9	2	2		1						21
	2	9	7	21	15	15	17	7	6	4	4	1	1	109

of qualitative analysis, and indicated that then Healy Picture Completion Test I should be solved by ten year old children, and that the Mare and Foal was usable for children under eight years of age. He also indicated that the Construction Puzzle A was useful for qualitative analysis of the methods of procedure in a task and the ability to profit by one's errors. The Ship Test, although used as a test for general intelligence, evidently does not adequately measure this as the correlation with Stanford is low and the distribution very irregular. It can be seen from table 4 that many cases as low as six years of age made the maximum score on the Ship Test, while those with

mental ages of twelve or above made low scores on the test. In other words, the test is poorly standardized at the lower and upper ends.

The results on these five tests indicate that age levels by these tests are indicators not of general intelligence level, but of some functions which require further analysis. A recent study of Construction Puzzles A and B, carried on at the Judge Baker foundation verified our results. The experimenters concluded that "In opposition to some earlier investigators we are forced

TABLE 4

Distribution on ship test

SHIP MENTAL AGE	STANFORD MENTAL AGE														TOTAL
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
12				3	2	3	1	6	2	6	2	1			26
11						1	1	3	1	1	1	1			9
10						1				1					2
9					2	1	7	4	1						15
8			1		3	1	2	3		1	2			1	14
7		1		3	2	2	1	1							10
6	1		2	2	9	4	1	1	1	1					22
5			2	5	4	2	6	1			1				21
	1	1	5	13	22	15	19	19	5	10	6	2	0	1	119

to conclude that neither Construction A nor Construction B, is an age level test" (17).

Healy Picture II has always been considered a test of special ability and we are inclined to agree with the belief. It shows a low correlation with Stanford and a fairly regular distribution. Healy indicated that this test was a test of apperception, considering apperception as essential to "any fair attempt to measure intelligence." Bronner, in further study of the test, characterized apperception as a special ability, but gave us no clear cut analysis of application of that ability.

Perry finds it to be a test of social adjustment. She found that many feeble-minded subjects who did well on this test were

more able to succeed socially than those who failed on it (18). It is suggested that this type of investigation be carried on with a larger number of cases to give more positive results.

The remaining two tests of the series, namely, Cancellation of A and Knox Cube Imitation, although giving low correlations with Stanford-Binet, have regular distributions. These results indicate that these tests of a special ability which needs further study. The results of the Cancellation test confirm those of Pyle, who found that the test has no close relation to general intelligence. Whether it is a test of attention still remains to be shown.

In the case of Knox Cube test, our results show that this test should not be used for the purpose of measuring what we call general intelligence. It is probably a test of a specialized ability, of visual rote memory, somewhat like that specialized ability found in repeating digits forward. The measuring and application of this remains to be found.

The writer wishes to express her appreciation of the help given by Irene Case Sherman in the collection and presentation of the data and by Jean Modell Edgerton in the collection of the data.

SUMMARY

1. Eleven commonly used performance tests were studied, namely, Pintner-Toops Directions, Seguin Form Board, Porteus Maze, Trabue Completion, Healy Picture Completion I, Healy Construction Puzzle A, Healy Mare and Foal, Glueck-Knox Ship, Healy Picture Completion II, Cancellation of A, Knox Cube Imitation.

2. The study was made by means of correlating the performance test mental ages with Stanford mental ages, and studying the distribution of mental ages.

3. The group studied was a selected group from the Institute for Juvenile Research, Chicago, Illinois, and had an intelligence level slightly below the normal group. Correlations were checked by repeated correlations on fifty new cases.

4. The correlations ranged from 0.41 ± 0.047 to 0.79 ± 0.001; see table 2 for complete list of correlations.

5. The Pintner-Toops Directions Test, the Seguin Form Board, the Porteus Maze, and the Trabue Completion Test gave high correlations with the Stanford-Binet and had normal distribution curves and, therefore, may be considered as partial substitutes for the Stanford-Binet mental age.

6. The Healy Picture Completion I, the Healy Construction Puzzle A, the Healy Mare and Foal, and the Glueck Ship correlated from 0.46 to 0.60 but showed such variations in distribution that they cannot, as standardized, be considered adequate mental age level tests.

7. The Healy Picture Completion II, the Cancellation of A, and the Knox Cube Imitation had correlations below 0.60 with regularity in the mental age distributions. We would suggest that these three are tests of special ability and are not mental age level tests comparable to the Stanford-Binet. Further analysis of these tests should be made to determine what abilities they do test and how they may be applied in clinical psychology.

REFERENCES

(1) HEALY, W. A.: Picture Completion Test. Psychol. Rev., 1914, xxi, 189–203.

(2) PINTNER, R., AND ANDERSON, M. M.: The Picture Completion Test. Educ. Psychol. Monog., Warwick and York, Baltimore.

(3) HEALY, W. A.: Picture Completion Test II. Jour. Applied. Psychol., 1921, v, 225–240.

(4) HEALY, W. A., AND FERNALD, G.: Tests for practical mental classification. Psychol. Monog., 1911, xiii, 53.

(5) PINTNER, R., AND PATERSON, D. G.: A Scale of Performance Tests. D. Appleton and Company, New York, 1917, p. 152.

(6) PINTNER, R., AND PATERSON, D. G.: A Scale of Performance Tests. D. Appleton and Company, New York, 1917, p. 152.

(7) SYLVESTER, R. H.: The Form Board Test. Psychol. Monog., 1913, xv, no. 4.

(8) KNOX, H. A.: A scale, based on the work at Ellis Island, for estimating mental defect. Jour. Amer. Med. Assoc., 1914, lxii, 741–747.

(9) PINTNER, R.: The standardization of Knox's Cube Test. Psychol. Rev., 1915, xxii, 377–401.

(10) KNOX, H. A.: A scale, based on the work at Ellis Island, for estimating mental defect. Jour. Amer. Med. Assoc., 1914, lxii, 741–747.

(11) PORTEUS, S. D.: The social rating scale. Training Sch. Bull., 1921, xviii, 33–39.

(12) PYLE, W. H.: A manual for the Mental and Physical Examination of School Children. University of Missouri Bulletin.

(13) PINTNER, R., AND TOOPS, H. A.: A Revised Directions Test. Jour. Educat. Psychol., 1918, ix, 123–142.

(14) WOODWORTH, R. S., AND WELLS, F. L.: Association Tests. Psychol. Monog., 1911, xiii, no. 57.

(15) TRABUE, M. R.: Completion-Test Language Scales. Teachers College, Columbia University, Contributions to Education, 1916, no. 77.

(16) JOHNSON, BUFORD, AND SCHRIEFER, LOUISE: Comparison of mental age scores obtained by performance tests, the Stanford Revision of the Binet-Simon Scale. Jour. Educat. Psychol., 1922, xiii, 408–417.

(17) LOWE, G. M., SHIMBERG, M. S., AND WOOD, M. W.: Further standardization of Construction Tests A and B. Jour. Applied Psychol., viii, no. 3., 324–338.

(18) PERRY, DORIS: Interpretations of the reactions of the feeble-minded on the Healy Picture Construction Test II. Jour. Delinquency, 1922, viii, 75–86.

A GROUP POINT SCALE FOR THE MEASUREMENT OF INTELLIGENCE

A Further Study in Method[1]

GRACE ARTHUR

Psychologist, Child Guidance Clinic, St. Paul, Minnesota

The present scale is an extension of the work reported in the Journal of Applied Psychology, June, 1919, by Arthur and Woodrow under the title, An Absolute Intelligence Scale: A Study in Method (1). The significant feature of this scale is that stressed in the earlier report, namely, the use of a formula for discriminative value. This formula reads

$$DV = \frac{Av_2 - Av_1}{\frac{\sigma_2 + \sigma_1}{2}}$$

In it, DV stands for discriminative value, Av_2, for the average of the higher age group under consideration, Av_1, for the average of the lower age group, and σ_2 and σ_1 represent the standard deviations of the distributions of the higher and lower groups respectively.

The use of this formula in the present scale is based upon two main assumptions, the first of which is that the value of any test consists in its ability to mark off any given group of a hierarchy from adjacent groups. In educational testing, the test in question must mark off a given grade from the grade above it and from the grade below it. In trade testing, the test must mark off the apprentice group from the novice on the one hand, and from the journeyman on the other, and so on. In intelligence testing, the discriminative value of any test

[1] The original thesis, of which the present report is an abbreviation, is on file in manuscript in the library of the University of Minnesota.

consists in its capacity to mark off any given group from the group of less ability on the one hand, and from the group of more ability on the other.

The determination of what shall be regarded as groups of less ability and groups of more ability involves the choice of a criterion of intelligence. This leads to the second assumption that the only objective criterion of intelligence available at the present time is the chronological age scale. Intelligence is here regarded as a function of a growth process, presumably that of the central nervous system. This growth process is, in turn, a function of chronological age. Increase in the ability to perform any task that involves the central nervous system is, then, regarded as being indirectly a function of chronological age, during the growth period.

The use of the discriminative value formula in the construction of the present scale, is therefore based upon the double assumption that any type of measurement of behavior involving the central nervous system, which serves to mark off one chronological age group from its adjacent age groups is a measurement of intelligence; and the more clear the demarcation, and the larger the number of age groups that are differentiated by it, the greater is its value as an intelligence test.

The first use to which the formula was put was to determine the discriminative values of the various tests used for all available age levels. The preliminary work furnished data for the calculation of these values. Table 1 is copied from the earlier report as it shows the difference in discriminative value between the opposites test and the cancellation test. Although the increase in crude score units from age to age is greater, for all but two age groups, for the cancellation test, its discriminative value is less than that of the opposites test at every age level. Because of its low discriminative value, the cancellation test was omitted from the revised scale. The digits test, although exhibiting slight increase in crude score from age to age, showed discriminative value at every age level considered, and on this basis was retained in the revised scale, although the plan had been to omit it.

TABLE 1

Comparison of point norms of easy opposites with three letter cancellation

AGE	EASY OPPOSITES				CANCELLATION			
	Original norms		Discriminative value	Point norms	Original norms		Discriminative value	Point norms
6.5	0	0		x	12.1	5.2		x
			0.9				0.8	
7.5	1.1	1.2		x + 0.9	16.0	4.7		x + 0.8
			11.1				0.3	
8.5	2.6	1.5		x + 2.0	17.2	3.9		x + 1.1
			0.9				0.8	
9.5	4.4	2.5		x + 2.9	20.6	5.0		x + 1.9
			0.34				0.0	
10.5	5.2	2.1		x + 3.2	20.6	4.9		x + 1.9
			0.65				0.5	
11.5	6.5	1.9		x + 3.9	23.0	5.1		x + 2.4
			0.40				0.1	
12.5	7.4	2.1		x + 4.3	23.7	5.7		x + 2.5
			0.55				0.3	
13.5	8.6	2.6		x + 4.9	25.6	5.6		x + 2.8

TABLE 2

Comparison of data from earlier and later studies for the Word Building Test

C.A.	EARLIER		LATER	
	Average	Points	Average	Points
6.5	0.4	x	0.4	x
7.5	2.3	1.1	1.5	0.84
8.5	4.3	1.9	3.2	1.7
9.5	6.3	2.5	4.2	2.1
10.5	7.4	2.8	6.3	2.8
11.5	9.8	3.5	7.3	3.2
12.5	10.3	3.6	8.7	3.5

It is interesting to note in table 2 that for populations of unequal ability, the discriminative values of a given test at a given level agree much more closely than do the crude score norms.

The second purpose which the discriminative value formula serves is the reduction of the crude score norms of the various tests to comparable units. The discriminative value of any test at any age level constitutes a point value expressed in terms of a unit of variability. These point values, expressed in the same kinds of units, standard deviation, for all the tests and for all available age levels, are combined by summation for the various age levels of the separate tests into separate scales, as shown in table 1 under "point norms." These point norms of the separate scales are in turn combined to form a single scale of all the tests used.

The third use to which the formula for discriminative value lends itself is to furnish a simple system of weighting that is independent of subjective evaluation of the tests to be weighted.

Referring again to table 1, we see that the discriminative values are larger for the opposites test than for the cancellation test at every age level. When these values are regarded as points, we see that a nine year score on the opposites test is credited with more points than a nine year score on the cancellation test. This means that if both tests had been included in the revised scale, success on the opposites test would have had a much greater part in determining the final rating of a subject than would success or failure on the cancellation test. This matter of weighting does not appear to affect average ratings in any great degree, but it does affect the validity of individual scores.

Results obtained with a scale constructed on the basis of the discriminative value formula are considered comparable with results obtained with the Binet scales, as the fundamental assumptions are the same for both types of scale: First, that the discriminative value of a test consists in its ability to mark off a given group from adjacent groups, and second, that the groups to be used for determining the discriminative value of

an intelligence test are furnished by the chronological age scale. The present method measures the discriminative capacity of a test in units of standard deviation. It was explained at some length in the earlier article that this value could be arrived at equally well by counting the number of cases of a group that fall between the average of a given group and the average of the next higher, or lower, group. The number of cases is changed into percentage of total number of cases, and this percentage is translated, by table, into sigma units. In the Binet method of scale construction, this same value may be arrived at by calculating the increase in per cent passing a given standard of achievement, between one group and the next.

This can be illustrated from the data of Terman, Lyman and Ordahl (2). Comparison of weights is passed by 50 per cent of the four-year-old subjects, and by 70 per cent of the five-year-old subjects. That is, 20 per cent more of the five-year-olds than of the four-year-olds are able to pass it. This means that if the capacity tested by this test is normally distributed, 20 per cent of the five-year-old group will have more of this ability than the average four-year-old, but less of it than the average five-year-old. This percentage constitutes the discriminative capacity of the test for these two groups, measured from the standpoint of the five-year-old group. If it had been larger, the discriminative capacity of the test would have been greater; if it had been smaller, the discriminative capacity would have been less.

It was explained at some length in the earlier report that the same portion of the area of the distribution surface that is represented by this 20 per cent, i.e., the portion of the area of the distribution surface for the five-year-old group that is included between the averages (50 percentiles) of the two groups, can also be determined by dividing the difference between the averages of the two groups by the standard deviation of the higher group and translating the result, by table, into percentage.

In the test used here for purposes of comparison the discriminative value of 20 per cent would determine automatically,

in the present scale, a weight of 0.53 points at this level. By the Binet method, this 20 per cent would be compared with the percentages obtained from other tests with the result that this test would either be discarded from the scale, i.e., given zero weight, or else retained and given equal weight with every other test at the same level. Where tests are of practically equal value the Binet method is satisfactory, but whenever it becomes necessary to use tests of very unequal value the present method would appear to be more sound in theory, more accurate mathematically and to lead to more satisfactory results in practice.

The chief differences between the present scale and the scales of the Binet type are that (*a*) the present scale is an amalgamation of separate test scales and not of tasks or portions of scales, (*b*) its results are expressed first in equal units, standard deviation, making possible a measurement of the increase of intelligence from age group to age group instead of being expressed in the variable units of mental age, (*c*) the discriminative capacity between two age groups of a test is affected by the standard deviation of both groups instead of by that of only the higher, and (*d*) the discriminative value of a test in the present scale is not only objectively determined, but objectively evaluated in its relation to the other tests of the scale. Of these differences, none are so fundamental as to interfere in any way with a similar use of results in determining intelligence quotients, or with a comparison of those quotients when obtained.

THE SCALE

The revised group point scale was constructed by applying the discriminative value formula to the crude score data obtained with the following tests:

Trabue Language Completion
Substitution
Comprehension
Opposites
Word Building
Digits

Each of these tests was given in two parallel forms. Scales B, C and L of the Trabue Language Completion tests were used; B, with Series I, and C and L with Series II. Two sets of the Woodworth and Wells Substitution blanks were ordered. Form I had the standard key at the top of the page. Form 2 had the digits of the key in reverse order. The symbols in the

TABLE 3

Opposite Test: Parallel lists of graded difficulty, based upon children's frequency tables

STIMULUS	PER CENT PASSING	POINT VALUE	STIMULUS	PER CENT PASSING	POINT VALUE
Yes	92.08	1.59	Rich	91.76	1.62
In	89.16	1.77	Young	88.75	1.79
New	85.16	1.96	East	85.50	1.94
North	81.83	2.09	Empty	81.00	2.12
Heavy	77.16	2.26	Dead	76.66	2.27
Sharp	72.00	2.42	Sour	71.08	2.44
Weak	66.00	2.59	Wild	65.91	2.59
Love	61.50	2.71	Less	61.16	2.72
Best	53.75	2.90	Buy	53.00	2.92
Over	48.16	3.04	Friend	45.75	3.10
Push	41.66	3.20	Broad	41.41	3.21
Much	35.00	3.38	Increase	34.75	3.38
Tender	29.33	3.54	Strength	27.25	3.60
False	23.33	3.72	Generous	22.75	3.74
Rude	19.25	3.86	Like	17.66	3.92
Stormy	14.08	4.07	Part	13.00	4.12
Bless	7.31	4.40	Silly	7.50	4.43
Frequently	4.58	4.67	Refined	4.16	4.71
Injurious	3.16	4.83	Permit	2.16	4.99
Strict	0.83	5.33	Simple	0.83	5.33

key of Form 2 were arranged in the following order: square, triangle, cross, star, circle.

The comprehension test had been constructed from three tests of the Kuhlmann-Binet Scale: drawing lines on squares according to directions, drawing squares on circles, and drawing triangles on squares. This test had demonstrated its value in the preliminary work. Form 2 was constructed by sub-

TABLE 4

Values of original norms when transformed into σ units or points

Av. = crude score norms; points = value in σ units.

C.A.	TRABUE SCORE			SUBSTITUTION AVERAGE			COMPREHENSION			OPPOSITES			WORD BUILDING			DIGITS			COMBINED CURVE
	Av.	Points	Scale	Av.	Points	Scale	Av.	Points	Scale	Av.	Points	Scale	Av.	Points	Scale	Av.	Points	Scale	Total points
6–0 to 6–11+	0.15	0.17	x	5.98	1.11	x	1.14	0.83	x	0.05	0.14	x	0.40		x	2.27		x	x
7–0 to 7–11+	2.75	1.45	1.45	15.34	1.27	1.27	2.29	0.72	0.72	1.25	1.11	1.11	1.52	0.84	0.84	3.82	1.14	1.14	6.53
8–0 to 8–11+	5.91	1.13	2.59	26.16	1.11	2.38	4.28	0.74	1.45	3.25	0.93	2.04	3.18	0.86	1.69	4.76	0.77	1.92	12.07
9–0 to 9–11+	7.62	0.56	3.15	33.05	0.66	3.04	6.12	0.50	1.95	4.54	0.50	2.54	4.21	0.43	2.12	5.08	0.32	2.24	15.04
10–0 to 10–11+	10.11	0.73	3.88	39.83	0.61	3.64	8.49	0.59	2.54	6.28	0.66	3.20	6.29	0.71	2.83	5.42	0.37	2.61	18.71
11–0 to 11–11+	11.41	0.34	4.23	45.37	0.49	4.13	11.00	0.55	3.08	7.30	0.38	3.58	7.31	0.31	3.15	5.74	0.34	2.95	21.12
12–0 to 12–11+	13.13	0.42	4.65	50.89	0.51	4.65	12.72	0.35	3.43	8.63	0.47	4.05	8.66	0.38	3.52	5.90	0.18	3.13	23.43
13–0 to 13–11+	15.79	0.62	5.27	55.31	0.40	5.05	14.23	0.33	3.77	10.55	0.70	4.75	11.24	0.64	4.17	6.18	0.32	3.44	26.45
14–0 to 14–11+	16.49	0.18	5.45	59.11	0.34	5.38	14.50	0.06	3.83	11.16	0.21	4.96	12.27	0.24	4.41	6.39	0.22	3.66	27.69
15–0 to 15–11+	17.92	0.32	5.77	62.63	0.31	5.69	15.51	0.24	4.06	12.20	0.34	5.30	13.98	0.39	4.80	6.48	0.09	3.75	29.37
16–0 to 16–11+	19.66	0.39	6.16	64.52	0.15	5.85	16.64	0.29	4.36	13.45	0.44	5.74	15.64	0.38	5.18	6.54	0.06	3.81	31.10
17–0 to 17–11+	20.19	0.11	6.27	67.07	0.20	6.04	17.20	0.16	4.52	14.00	0.22	5.96	16.80	0.25	5.43	6.58	0.04	3.85	32.07
18–0 to 18–11+	22.98	0.60	6.88	72.83	0.53	6.57	19.27	0.80	5.32	15.62	0.78	6.74	18.20	0.33	5.76	6.87	0.33	4.18	35.56

stituting "up" for "down," and "right" for "left," and vice versa, in the directions of Form 1.

Work with the opposites tests of the earlier scale had shown the need of standardization for children of the separate words of the lists used. Responses were obtained from 600 school children of grades II to VIII, inclusive, for 96 stimulus words. Two graded lists of approximately equal difficulty were prepared on the basis of these results and were incorporated in the revised scale. These lists are presented in table 3 with the per cent of successful response to each stimulus word, and the point value of each in units of standard deviation (3).

The two forms of the word building test, a, e, o, b, m, t and e, a, i, r, l, p were given and scored according to the directions given by Whipple. This test and the digits test were the only two tests of the preliminary series to be used without change in the revised scale.

Forms 1 and 2 of these six tests were assembled into Series I and II. The parallel forms were used, not for the purpose of supplying alternative tests, as might be supposed, but to increase the reliability of individual scores. The two series were presented approximately one month apart. Results from the two forms were averaged for each test.

Crude scores were obtained with these tests for 2517 pupils of grades I to XII inclusive. The discriminative value of each test at each age level was calculated. These values were combined by summation for each test, as shown in table 4. The cumulative value constituted the point norm corresponding to the crude score norm for each test at each age level. Between the year norms point values were calculated for all possible scores falling between the crude score norms. It was assumed that the standard deviation of the 18.5-year group would remain constant for all higher age groups, and point values were calculated upon this basis for scores falling above the 18.5-year norm. It was also assumed that the standard deviation of the 6.5-year age group would remain constant for all lower age groups, and upon this basis point values were calculated for all scores falling below the 6.5-year norm. There is, of

course, an element of error in both these assumptions, but it is, presumably, not serious inasmuch as the scale was given to children of 6.5 to 18.5 years, inclusive, and only a scattering handful of cases fell outside these limits. By means of the separate scales so constructed, individual crude scores for the separate tests were converted into point scores for those tests.

A combined scale was constructed by adding the point norms for the separate tests at each age level. Values were interpolated between the year norms of the combined point scale to correspond with months of mental age.

The procedure, then, in finding the rating of a given subject was (*a*) to find the crude score on each form of each separate test, (*b*) to average the scores of the two forms of each test, (*c*) to translate the average crude score into point score for each test, (*d*) to add the point scores of the separate tests, (*e*) to translate point scores, by table, into years and months of mental age, and hence to IQ.

Data obtained from foreign pupils with less than one year of experience in an English speaking school were disregarded. The pupils whose scores were considered valid constitute a group that is not unselected in one sense, as a very large percentage of them were of foreign birth or parentage. But they were as nearly unselected a group as one could reasonably hope to obtain for the ages 6 to 15 inclusive, in the sense that all the individuals of these ages in the given school district at the time the tests were administered were present and were represented by test scores. The only cases of school age not represented by test scores would be those under detention in state institutions and those under quarantine for some infectious disease. Of the former, there were three cases in the state school for the feebleminded, and one in the state school for the deaf; and of the latter, there were several families under quarantine for scarlet fever.

These children were not "test wise." For most of them, the intelligence test situation was wholly new.

RESULTS

In the preliminary study were published intelligence quotients for 51 cases tested individually by the Kuhlmann 1917 Revision of the Binet-Simon scale. The uncorrected coefficient of correlation between the two sets of quotients was found to be 0.78 (Bravais-Pearson). This we considered surprisingly high at the time, in view of the greater accuracy of individual tests as compared with group tests, and considering the fact that some tests which were included in the group scale at that time had been shown to possess very low discriminative value. To quote directly from the earlier report: "We incline to attribute a considerable portion of the apparent success of the present tests to the method we have used, and to believe that inadequacy of the tests and norms is compensated by the advantage inherent in the methodological principles we have followed."

In the present study 275 cases were available for a similar comparison. Among these were the pupils of the five Grade II classes of a large grade school, besides cases referred for testing prior to grade placement, and "problem" cases. In figure 1 is given the scatter-graph showing the correlation between these 275 individual intelligence quotients obtained by the Kuhlmann-Binet Scale (1922) and the group intelligence quotients of the present group scale for the same subjects. For these, the uncorrected coefficient of correlation is 0.917 (Bravais-Pearson), or, discarding third place decimals, 0.92 ± 0.006. The correlation ratio also was determined. This, too, was found to be 0.92. The regression, therefore, is linear, according to Blakeman's criterion (4).

As the two sets of quotients appeared to run so closely parallel to each other, it seemed worth while to calculate their probable error, also. For this the Nygaard formula was used (5). This formula is based upon the average difference between the two sets of values. It yields a value that is independent of the form of the distribution curve, and that is not affected by the range of cases. The probable error was found to be 2.60. This means

that intelligence quotients obtained by this scale at a given date agree more closely with Kuhlmann-Binet intelligence quotients obtained over a period of three years, than those for the Stanford-Binet agree with re-tests made with that same scale. Quoting directly from Terman: "Retests have been given to 315 children in the vicinity of Stanford University. To 46

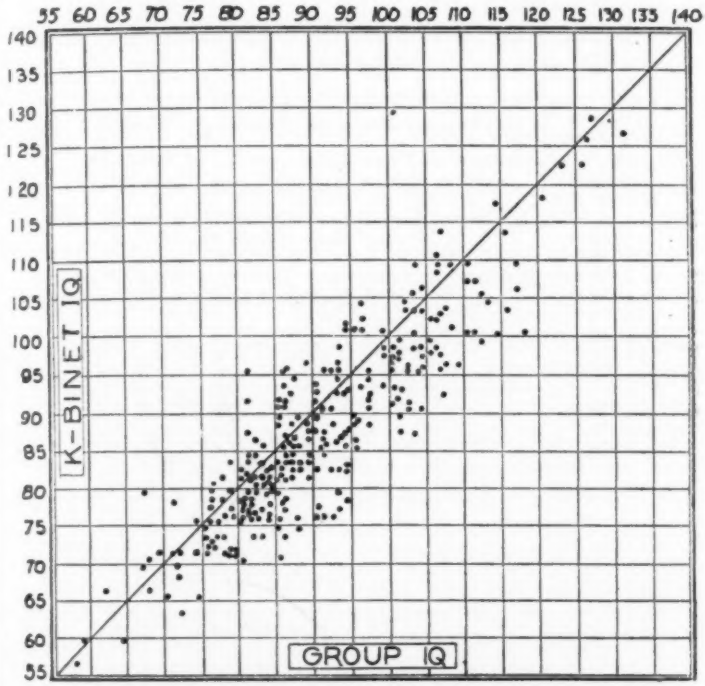


FIG. 1. SCATTERGRAPH SHOWING CORRELATION BETWEEN KUHLMAN-BINET IQ'S AND GROUP POINT SCALE IQ'S

of these children three or more tests have been given. In case of a child tested several times, each test has been compared with each of the others. . . . . This gives in all 435 comparisons . . . . The probable error of a prediction based on a first test is 4.5 points in terms of IQ." (6). Terman does not state what formula was used in calculating this value.

TABLE 8

Showing distribution by mental age of each chronological age group

C.A.	6–6	6–6 TO 6–11+	7–0 TO 7–11+	8–0 TO 8–11+	9–0 TO 9–11+	10–0 TO 10–11+	11–0 TO 11–11+	12–0 TO 12–11+	13–0 TO 13–11+	14–0 TO 14–11+	15–0 TO 15–11+	16–0 TO 16–11+	17–0 TO 17–11+	18–0 TO 18–11+	18–6	TOTAL CASES
22–0 to 22–11+															1	1
21–0 to 21–11+											1		2			3
20–0 to 20–11+												1			2	3
19–0 to 19–11+									1		1		2	2	7	13
18–0 to 18–11+											3	1	6	4	14	28
17–0 to 17–11+								5	5	2	2	8	9	7	20	58
16–0 to 16–11+							4	6	6	7	8	5	17	6	17	76
15–0 to 15–11+						4	9	15	10	11	20	16	19	12	15	131
14–0 to 14–11+					4	9	11	34	26	22	16	7	22	9	13	173
13–0 to 12–11+			1	4	5	16	13	47	21	31	29	12	14	6	3	202
12–0 to 12–11+				4	16	36	33	51	41	20	13	6	1	4		225
11–0 to 11–11+		1	2	7	49	56	52	54	29	7	7	2				266
10–0 to 10–11+			5	32	53	58	51	36	10	4						249
9–0 to 9–11+	2	2	37	69	82	57	20	11	2							282
8–0 to 8–11+	6	12	64	96	55	18	8									259
7–0 to 7–11+	43	57	103	66	16	1										286
6–0 to 6–11+	149	85	18	2												254
5–0 to 5–11+	5	3														8
Cases.....	205	160	230	280	280	255	201	259	151	104	100	58	92	50	92	2,517

Dr. Kuhlmann insists that the comparison of the probable error of measurement between the present scale and the Kuhlmann-Binet with that of the Stanford-Binet is unfair unless the fact is stressed that the present group scale consists of two parallel series of tests.

In the preliminary study, mental ages were recalculated from the group scale data for the 52 subjects for whom Kuhlmann-Binet scores had been obtained, using the year method. IQ's were calculated and compared with those obtained by the point method. The difference between the two series of IQ's ranged from 0 to 9, with an average difference of 3.4 points. Upon the basis of the average amount of error, the extra labor involved in the use of the point scale is evidently not worth while. From the standpoint of individual diagnosis, however, nine points of IQ constitute a significant error, and one to be avoided if possible. Point scale IQ's tended to run higher than those obtained by the year method, and to agree more closely with the results of the Kuhlmann-Binet scale.

In table 5 is given the distribution of cases of each chronological age group by mental age. It is largely self-explanatory. The bunching of cases at the levels <6.5 and >18.5 is of course due to the lack of differentiation in terms of mental age of cases falling within those groups. The cases of >18.5 are differentiated in terms of points, but a large percentage of those in the <6.5 group are not.

It will be noted that below the twelve year level there is a definite downward skew to the distribution. This is due to a tendency, inherent in a system of weighting based upon successive averages and standard deviations, for the subjects in the lower age groups to be penalized for falling below their appropriate age norms in some tests to a greater degree than they are rewarded for surpassing those norms in other tests. This can be illustrated by the hypothetical case of a 10.5-year-old boy whose scores range from the 6.5-year norm in Digits, to the 12.5-year norm in Opposites. If his Substitution score falls at the 9.5-year norm, his Comprehension and Word Building scores fall at the 10.5-year norms for those tests, and his

Language Completion score falls at the eleven-year norm, we find that he has been penalized 0.69 point for falling below the 10.5-year norm in Digits, and 0.60 point for falling below in Substitution, while he is rewarded 0.35 for exceeding the 10.5-year norm in Language Completion and 0.85 for exceeding it by two years in Opposites. Comparing total rewards with total penalties, we find that he has been penalized 0.09 point more than he has been rewarded, and that in consequence his total point score falls 0.09 point short of the 10.5-point norm, giving him a mental age of one month less, or ten years and five months. As stated above, this is the inevitable result of the tendency toward decrease in the difference between the successive averages as we ascend the chronological age scale, and of the tendency toward increase in the successive standard deviations. Both of these tendencies operate to decrease the size of the quotient yielded by the formula for discriminative value, at successive age levels, and consequently to decrease the added increment of points at those levels. This results in a definite, though slight, downward skew in the distributions of all age groups unless offset by counter-tendencies resulting from the selection of groups above the age of compulsory school attendance.

This slight downward skew was detected also in the distributions of the preliminary study. It is what should be expected when we use units of equal significance in measuring the ability of groups in which intelligence, in terms of years of mental age, is distributed according to the normal probability curve. For if the distribution of intelligence is normal when measured by the unequal units of the mental age scale, it must, of necessity, be skewed when measured by the equal units of the present point scale.

At the higher end of the scale there is an upward skew due, probably, to the selection operating above the age of compulsory school attendance. This selection results in a tendency for the differences between successive averages to increase, and for the successive standard deviations to decrease, with a consequent increase in the size of the quotient yielded by the

formula for discriminative value. Success at the higher levels, therefore, tends to be over-rewarded and failure tends to be under-penalized.

The downward skew of the lower age groups, and the upward skew of these higher groups can be corrected by the use of average point scores as norms. This method yields a slightly more accurate rating. The scale as used in the present study is more flexible than a scale constructed of average point scores however, in that tests can be added or omitted at will without re-standardizing the scale as a whole.

SUMMARY

The significant feature of the present scale is the use of a formula for discriminative value. This formula reads

$$DV = \frac{Av_2 - Av_1}{\frac{\sigma_2 + \sigma_1}{2}}$$

In it, DV stands for discriminative value, Av_2, for the average of the higher age group under consideration, Av_1, for the average of the lower age group, and σ_2 and σ_1 represent the standard deviations of the distributions of the higher and lower groups respectively.

The first use of the formula was to determine the discriminative values of the various tests used, for all available age levels.

The second purpose which the formula serves is the reduction of the crude score norms of the various tests to comparable units.

The third use of the discriminative value formula is to furnish a simple system of weighting that is independent of subjective evaluation of the tests to be weighted.

Results obtained with a scale constructed on the basis of this formula are considered comparable with results obtained by the use of the Binet scales, as the fundamental assumptions are the same for both types of scales: first, that the discriminative value of a test consists in its ability to mark off a given group from adjacent groups, and second, that the groups to

be used for determining the discriminative value of an intelligence test is furnished by the chronological age scale.

The revised group point scale was constructed by applying the discriminative value formula to the crude score data obtained with the following tests: Trabue Language Completion, Substitution, Comprehension, Opposites, Word Building and Digits. Each of these tests was given in two parallel forms. Crude scores were obtained with these tests for 2517 pupils of grades I to XII, inclusive. These were translated by table into point scores, and thence into mental age and IQ

Of these pupils, 275 had been tested individually on the Kuhlmann-Binet scale. Comparison of the group IQ with the individual IQ yields an uncorrected coefficient of correlation of 0.92, ± 0.006 (Bravais-Pearson). The correlation ratio also was calculated, and the regression was found to be linear, according to Blakeman's criterion. The probable error of measurement for the two sets of quotients, according to Nygaard's formula, was found to be 2.60.

REFERENCES

(1) Arthur and Woodrow 1919 An absolute intelligence scale: A study in method. Jour. Applied Psychol., vol. iii, 118–137.

(2) Terman, Lewis M., et al. 1917 The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence, p. 166.

(3) Arthur, Grace 1922 A standardization of certain opposites for children of grade school age. Jour. Educat. Psychol., 483–495.

(4) Blakeman, J. Biometrika, vol. 4, 349–350.

(5) Nygaard, P. H. 1923 Advantages of the P.E. of measurement as a criterion of the reliability of a test or scale. Jour. Educat. Psychol., vol. 14, 407–413.

(6) Terman, Lewis M. 1917 Intelligence of School Children, pp. 138–142. Houghton Mifflin Company, Boston.

SLANG AS AN INDICATION OF CHARACTER

GLADYS C. SCHWESINGER

Teachers College, Columbia University

The research here reported is an exploratory attempt in the field of character detection and measurement. The aim was to find an instrument which would differentiate delinquent children from non-delinquents. Other factors of age, mental age and intelligence quotient, of social and economic status, and even to some degree of nationality and race were to be kept constant. It was planned to construct tests along a new and original line and to apply these to two opposed groups of children, groups which, at least in so far as the experimenter could determine, differed only in the matter of moral character. A comparative study of the results was to be made in which the separate elements would be examined for whatever significance each might reveal. All such elements which survived the application of Yule's formula for determining the significance of differences were to be retained and thrown together into a new composite which it was hoped would bear up under later verification processes.

If a test could be evolved which would objectively pick out delinquents from non-delinquents, which would aid in the detection of potential delinquents, which would indicate to the examiner in some quantitative way the degree of delinquency of the examinee, the attempt would be more than justified. The main previous effort in this direction has come from Whittier, California. The Whittier scale, doubtless familiar to all interested workers, has been derived from careful scientific analyses of records of crime at the Whittier State Home for Boys. It consists of a series of offences listed on a chart, horizontally, according to the character or nature of the misdeed, from the least serious type to the most serious, from such light offences

as truancy to such grave crimes as assault and battery and murder. Vertically, each kind of offence is graded for degree of participation, for example, from the first offence to an habitual one, or from the theft of some article of but little value to one of great worth.

A child's record is examined and his misdemeanors plotted on this charted list. His demerits are then totaled and he is given a score. The scheme works adequately when an intimate history of the case over a wide period of time is available. But often a child may be brought to court about whom relatively little is known except for the offence which precipitated him there. He may be a comparatively "good" boy, who having fallen but once from grace, fell hard. If his offence, and it may have been his first, is for some serious charge such as grand larceny, his score will be high. Rated side by side with some youngster who was constantly in trouble for petty thieving, truancy, and the like, the newcomer, by virtue of his Whittier index stands a chance of being characterized as the more delinquent. The scale aims, but often fails, to give an inclusive quantitative picture.

One of the problems of state institutions is how and when to determine parole. Many entrants are more or less parolable from the very beginning. Commitment within the walls of the institution is salutary lesson enough; continued residence is not necessary and for various reasons undesirable both for the boy and for the State. A test, therefore, which would sound the boy's behavior level, which would tap his conduct tendencies, both past and prospective, would assist materially in determining the moral status of the boy.

Prognosis might advisedly date even further back. If potential delinquents could be selected, possibly segregated, certainly supervised, before being given time to launch on their career of crime, many of the hardships of conduct maladjustment might be avoided. Preventive, therapeutic and educational forces could be brought to play on an incipient case with much greater effect than on the hardened reprobate of several months or years standing.

The only defence offered for the character and type of test material here chosen was an effort on the part of the author of the tests to tackle the problem from what might be the delinquent's end of the game. Ordinary tests usually contain words, ideas, and situations which may well be supposed to be within the reach and ken of every normal boy. If, in the devising of new tests, something could be assembled which might tap the experience of the delinquent rather more than the conformist greater possibilities of differentiating results might be hoped for.

The question of knowledge of crime and of life behind the scenes of conventional standards, and conduct in accordance with or against those standards, need not be discussed here. Suffice it to say that the tests evolved represent an effort to reach not only the boy's knowledge of what is socially undesirable but also his tendency to put the same into practice.

True situations calling for moral reactions, or their opposite, to be effective, would have to be concrete. No doubt many tasks could be devised where the purpose of the test situation is cleverly hidden from the subject, where he would be compelled to make a choice of reaction and be rated accordingly, but due to the pressure of time and circumstance, general administrative difficulties involved in giving and scoring individual tests, the scope of possibilities was here confined to pencil and paper and to group work. If the sum total of man's reactions can be catalogued into the threefold, thinking, feeling and doing, and if, as has already been hinted, what one thinks and feels is not necessarily consistent with what one actually does, the experimenter is compelled to confine his attention to the last named type of behavior. If our overt reaction were to be subdivided we might make the line fall between what one does and what one says. We tend to associate bad language with delinquent behavior, but is one an index of the other? How bad, or rather how good, can language be before its use becomes significant of the conduct of its spokesman? In other words, does the language used by the delinquent differ materially from that used by the so-called good boy? Does the delinquent have a wider knowledge of certain colloquial, slang and

vulgar phrases? Does he evince greater familiarity with the terminology of the street or of the professional "dip?"

No effort was made to test for actual swearing, nor for practical reasons, was the full gamut of youth's misconduct sounded, but a great collection of slang phrases and street language was gathered and presented to the subjects in four different groupings: (1) definitions of single words, (2) definitions of phrases, (3) Same-Opposites, (4) classification of slang knowledge. For the first two groups the multiple choice method of response was offered, the selection of answer to be made from one of four alternatives: (*a*) a true meaning of the word or phrase expressed in good English, (*b*) a meaning not true, but plausible, in good English, (*c*) a slang meaning, but true, (*d*) a slang meaning but false. Interspersed with the slang words were some good literary English words, some of which were a little unusual, chosen from collections like the Thorndike Word Knowledge Tests. Interspersed with the slang phrases were some good English phrases or proverbs chosen from collections like those in Kohs' list. In the Same-Opposites test, slang words were paired sometimes with good English words, sometimes with slang words. Checks of paired literary words were scattered throughout. In group 4, the candidate was required to identify his word with the activity or type of delinquency with which it was ordinarily associated. The words were all indicative of the activities of the lower tenth of the social order, and were mostly slang, street language and in some cases professional "crook" terms.

Attention is called to the fact that scattered throughout the four groups, provision was made for a fairly liberal supply of ordinary English words which could be used not only as a rough estimate of the subject's general word knowledge, but also as a check against a possible tendency to react, or to fail to react to the test elements because of strangeness or distaste.[1]

With the coöperation of Mr. G. M. Willson, Assistant Psychologist at the New Jersey State Home for Boys in Jamesburg,

[1] A copy of the tests may be found at the end of this article.

to whom grateful acknowledgement is here made, it was arranged to give the tests to the boys in that institution. A previous preliminary try-out had revealed the undesirability of attempting to test boys under ten years mentally, whether delinquent or normal in social behavior. At Jamesburg, the files were carefully combed and only those boys were selected who were bright enough to understand and follow directions on a group test. In June a visit was made to the institution and test results of fifteen boys were analyzed and compared with the results of fifteen control boys from a small town in New Jersey. These results were significant enough to justify the further pursuit for data. In August the number was completed to one hundred cases, ranging in chronological age from ten to eighteen years with a median of 14.8. In mental age the range was from nine to eighteen, the median falling at 12.78. Intelligence quotients spread from 67 to 144, with a median of 88.25. This group was later closely matched for distribution of C.A.'s and M.A.'s, and for social and economic status from four public schools of a town in New Jersey.

Parental nationalities represented in each group were American, Italian, Colored, Hungarian, Irish, German, French. Some Russians and Poles were included only in the delinquent group, and some English, Danish and Swedish were in the control but not in the delinquent group. Catholics, Protestants and Jews were proportioned rather evenly in both groups. Sample occupations to be found in each group include those of laborer, carpenter, machinist, painter, factory-hand, truck driver, fire department and police service, stableman, printer, tailor, chauffeur, contractor, gardener and the like.

A side study of various methods of scoring the tests was made with the hope that by penalizing here or there a wider differentiation might be obtained. But this was not realized. It makes little or no difference whether a point is given for using the slang rather than the literary word in the multiple choice responses, although analysis of the results showed a slightly greater tendency on the part of the delinquents to express themselves in slang than for the controls. No material differ-

ence was apparent when penalties were assigned for giving the wrong answer, except that in the lower cases it became a question of handling minus scores. For this reason it was decided to keep to the simple scoring technique of assigning a point to each correct response whether or not that response was a literary or a slang word.

The results of the tests, when scored and tabulated for averages and dispersions show a fair and consistent differentiating power. Taken separately, these tests, with the exception of the fourth, under the caption of "Classifications" do not show very large P.E. differences, but because this difference is always on the one side, the four combined produced differential results which were acceptably significant, yielding a total difference of 7.1 P.E., 7.6 P.E., or 6.1 P.E., depending on the way the responses were scored.

Reference has already been made to the non-slang check words and phrases which were scattered throughout the four pages of test material. These were scored separately and the results are nothing short of astonishing in that the averages and distributions are almost identical for the two groups. On the slang material, 76 per cent of the delinquents reached or exceeded the median of the unselected children. On the literary words, exactly 50 per cent of the delinquents reached or exceeded the median score of the control group.

Considering that the slang battery revealed a substantial differentiating power and that the vocabulary of non-slang words failed to show such results, it was felt that here might be something of value which was worthy of further research. Consequently the writer undertook a careful and exhaustive analysis of all the material gathered. Every word and phrase in the slang battery was tabulated and recorded for its frequency of response. They by applying the formula advocated by Yule[2] the significance of each difference was determined. Where a sigma value of 2.5 and in some cases 2 was revealed that element was retained. The rest were thrown out.

Those elements which were found to have differentiating

[2] "Introduction to the Theory of Statistics," p. 269, G. Udney Yule.

power were reassembled into a new composite together with many interspersed neutral elements which were to be disregarded in the scoring, the object of their inclusion being to keep some of the easier and more familiar words and phrases on the page and thus produce a more normal emotional reaction from the subject than would be experienced were he to be confronted suddenly with only the extreme elements. The words retained represented in every case a greater degree of response from the delinquent than from the non-delinquent group, many of these being crime words, or at least words representing an acquaintance with various forms of social misconduct.

A probation school in Brooklyn was enlisted in which to try out the revision. Four classes of boys were tested by the National Intelligence Test and all falling below a mental age of ten years were discarded. The fifty cases remaining were matched from a public school, also in Brooklyn, the boys being of the same social level and in the corresponding school grades. The groups were equated for averages and variabilities of mental age and intelligence quotient.

The results of this testing were disappointing. No appreciable difference was found between the responses of the two sets of boys. The averages and dispersions of the two groups were about even.

Possible explanations are always in order. Since the problem of establishing a criterion and validating character tests has been at all times a great one, and since the writer failed finally to get differentiation by the method of comparing averages of two supposedly distinct groups, the question arises whether or not this failure is due to the fact that there may be no real psychological difference between the two groups, or that there may be a weakness in the method of arriving at conclusions.

The children in the probation school may or may not have been afflicted with the same sort of conduct disturbance which caused the Jamesburg boys to be committed. If there is any character weakness common to both we may feel reasonably safe in saying that the State Institution residents have it in greater degree, since they could no longer be retained in an

ordinary social environment. The records at Jamesburg read some thing like this: breaking, entering and larceny, vagrancy, incorrigibility, malicious mischief, persistent truancy, grand larceny, shooting teacher, assault, and the like. Many cases offer combinations of several of these and other offences. The Brooklyn Probation School records offences of: school incorrigibility, disobedience, truancy, smoking, obscene language, laziness, stealing, pugnacity, sullenness, and the like.

Whether or not these two groups of boys represent steps on a scale of misconduct, the fact remains that the tests did not pick out the school incorrigibles from the school boys who were not rated as incorrigible. Part of the value of the tests, it was hoped, would lie in their prognosis of potential delinquents and in this apparently they fall short.

Of course, it may be that the probation group does not cover the delinquents of the future. Certainly the boys cooperated well and obediently in taking the tests. Discipline within the school room was rigid and exacting. Four teachers were present and there was no sign whatever of insubordination on the part of any pupil. Perhaps the reformative influence had already set in!

Again, it may be possible that the control group contained as many "bad" boys as were to be found in the delinquent group, but that because of a certain cunning or shrewdness, they had failed to be apprehended, or in other words, that the delinquents were not really selected and the controls unselected; that the two groups were comparable in every way, including the trait which the experimenter was trying to measure, and hence rendering differentiating results impossible.

The question of method of arriving at conclusions becomes a vital issue. The method of comparing contrasted groups presupposes a knowledge of the character of the groups studied. The fallacy may lie here. It might be better in future to confine attention to those children about whom definite and detailed and complete historical information is available, to correlate test scores with the evidence at hand, whether that be objective or subjective. In the case of established and recorded

delinquents, a Delinquency Rating could be worked out consistent with the historical record. Where official papers are not available, the evidence could be some approved tangible, concrete and factual rating scheme, rated by people who are in a position to make judgments. Cady has shown the importance of close acquaintance and confidence on the part of the judge.

It might be advisable to study the relationship of judgments of this latter type with ratings on the former scheme, judges to be ignorant of the facts set down in official records. It might cast some light on whether or not school incorrigibility, which is probably the basis for a teacher's judgment is a different thing from the child's anti-social behavior which brings him up for commitment through the courts. Enlightenment here might cast some light on why different results were obtained in a State Institution group from those obtained in a public school probation group.

For whatever they may be worth to others the slang tests of the first survey are appended. While by no means exhaustive of slang in current usage, they represent an effort to cover those words and phrases which might fall within the experience of an ordinary boy or girl who has been exposed to an environment which makes use of slang as an avenue of expression. If some future investigator is interested in the question of verbal tests as a measure of intelligence, involving the issue of unfair advantage which they give to the child of superior environment, he might do well to take these "language" tests which represent a lower social order and make a comparative study of different social levels, casting aside the factor of conduct and confining his attention solely to the question of knowledge.

APPENDIX

DEFINITIONS

Look at this sample:

1. SNITCHING means . . . . sewing, squeeling, stealing, telling

SNITCHING means "squeeling," so a line is drawn under "squeeling."

Now look at sample 2:

2. CRACKED means . . . . broken, looney, cranky, crooked

Which of the last four words best explains the meaning of CRACKED? Either "broken" or "looney" will do, so draw a line under either one, but only under one.

Now on the rest of the page are a great number of exercises. Remember, underline only *one word* on each line, that word which will give the best meaning. Begin at the top and work all the way down the list until the time is called.

SIMP	means dumbell, simpleton, sink, scab
IVORY DOME	means cupola, bonehead, elephant's tusk, foolish
IRON MEN	means Crusaders, dollars, bones, rubes
SHEENEY	means Silky, Kike, Jips, Jew
TO BE STEWED	means drunk, cooked, tanked, batty
COTTON	means cloth, hut, flour, herd
THIS JOINT	means this house, knee, plunk, den
SQUEELING	means telling, yelling, whooping, tattling
VAMOSE	means beat it, car, go away, skidaddle
HOOTCH	means booze, beer, hatch, hoodlum
DASH	means rush, smear, lack, fasten
DEVOUT	means lawless, pious, devotion, formal
DIP	means cannon, vag, dive, thief
ELBOW	means collar, detective, cinch, bone
TIGHT-WAD	means booze-artist, narrow, stingy, close-fisted
FOBBER	means fob-worker, pick-pocket, watch-fob, mik
NUTS	means foolish, dippy, fruit, mut
ABILITY	means diligence, speed, power, dignity
LIGHT-FINGERED	means fairy-like, thievish, three-fingered, pussyfooted
CARGO	means load, hem, draught, vehicle
DUCK	means fowl, hide, beat it, pippin
WARM BABY	affectionate, woollen clothes, Hot Tamales, boob
DUKES	means hands, mits, aristocrats, bucks

SHYSTER means bum mouthpiece, crook, bad lawyer, shy
CLERGYMAN means grandsire, employee, beadle, minister

PHRASES

Below are a great many expressions. Four answers are given after each expression. You are to look at the answers carefully; then make a cross in the space before the best answer to each expression, the one, which to you, best explains its meaning, as in the sample.

Sample: What do we mean by "Shake a leg"?
—1. Exercise yourself
—2. Won't you shimmy for me?
—3. Act quickly
—4. Get a move on

Begin with No. 1 and keep on doing all the others until time is called.

LET GEORGE DO IT means
— Pass the buck
— Avoid accepting responsibility
— Let George try if he wants to
— To get away with it

HE HAS A SCREW LOOSE means
— He needs to be mended
— He isn't quite sane
— He is a hunk o' cheese
— He is dippy

COME THRO' WITH THE GOODS means
— Deliver them to our house
— Cough up
— Give me your money
— Snow again

DIG A LITTLE DEEPER means
— Contribute
— Fish up
— Make a bigger hole
— Call him down

THEY HANDED ME A LEMON means
— They gave me a yellow fruit
— They slipped one over on me
— They satisfied my thirst
— They handed me the straight dope

COME OUT OF YOUR TRANCE means
— Step on the gas
— Wake up
— Perhaps you're only dreaming
— Go up in the air

DON'T CRY OVER SPILT MILK means
— A watched pot never boils
— The dish ran away with the spoon
— Don't cry over little things
— Don't worry about things you can't help

FORGET IT means
— Shut up
— Don't mention it any more
— Try not to remember it
— Fall for it

SHE'S PULLING YOUR LEG means
— She's making you taller
— She's got your number
— She's trying to get something
— She's working you for something

TELL IT TO SWEENEY means
— Pass on the news to Sweeney
— I don't believe you
— Don't kid the goldfish
— Get wise

GO TAKE THE AIR means
— You look a little faint
— Slow down
— It's time to leave
— Get off me lip

THEY'RE STUFFING HER means
— They're hitting the ceiling
— They're feeding her a lot
— They're kidding her
— They're deceiving her

BIRDS OF A FEATHER FLOCK TOGETHER means
— Hunger is a dangerous companion
— Pretty maids all in a row
— Good company makes short miles
— You are known by the company you keep

THIN IN THE UPPER CRUST means
— Haven't much brain power
— The pie isn't baked enough
— He's cracked
— To bark up the wrong tree

TO GET ALL THE DOPE means
— Collect the information
— To be a cocaine (drug) victim

— Round up the goods
— To land in the jug

HE LED ME A MERRY CHASE means
— We had lots of fun together
— He strung me along
— He is on the level
— He kept me guessing

WHERE DO YOU HANG OUT? means
— Where is your clothes' line?
— Where do you live?
— Where do you bunk?
— Are you talking through your hat?

HOW DO YOU GET THAT WAY? means
— What made you reach that conclusion?
— What have you done to yourself?
— What's eating you?
— Your face, only a mother could love

GIVE 'EM THE RAZZ means
— Make yourself solid with them
— Rub it in
— Turn on the victrola
— Tease them

TAKE THE MAGGOTS OFF A BLIND MAN'S EYE means
— To be close-fisted
— To be kind to an unfortunate man
— To be stingy
— To freeze out

HE IS GOING TO TAKE A RIDE means
— He is going to jail
— He is going out in his car
— They will put him in the cooler
— He has to ride the goat

ROME WAS NOT BUILT IN A DAY means
— I've been to London to see the queen
— In time, one man's loss is another's profit
— Better a day too soon than a minute too late
— It takes time to produce anything of value

NOW YOU SPRING SOMETHING means
— Its up to you to wobble now
— Don't monkey with the buzz-saw
— Its your turn to tell a story
— Press the button

HE'S NUTS means
— He's loco

— He is very foolish
— He likes to go nutting
— He'll pass in his checks

SPEEL 'EM OFF means
— To say them one at a time
— To take the skins off
— Let her shoot now
— Cut out the rough stuff

I GOT THE CAN TIED TO ME means
— They gave me the grand bounce
— There's a tin fastened to my belt
— I didn't do her nothing
— I have been dismissed

BAG YOUR HEAD means
— Keep quiet
— Shut up
— Put your cap on
— Lay down and die

PIPE DOWN means
— Keep still
— Lay your pipe down
— Put on the brakes
— Can the comedy

HE'S NIGGIN' means
— He's a negro
— He is cheating
— That's slight of hand stuff
— He's pulling the wires

MAKE HAY WHILE THE SUN SHINES means
— Make friends, they increase your joy
— He who sings drives away sorrow
— Make the most of your opportunities
— The sheeps in the meadow, the cows in the corn

CUT IT OUT means
— Don't do it any more
— Sneeze out of it
— Chuck it
— Use a knife

BY THE SKIN OF HIS TEETH means
— Just by a narrow margin
— His teeth are decaying
— It's a close shave
— Hitting the high spots

BEAT IT, BUM means
— Shake the rug
— Go away
— Make tracks
— Give him a knock-down

NOBODY HOME means
— The family has gone out
— He is feeble-minded
— They froze him out
— He is loose on top

TO GIVE HIM THE SACK means
— To fire him
— To give him a bag
— To make him give up his job
— To hand him one

IT'S THE CAT'S MEOW means
— He is most important
— He is mewing like a cat
— He's the cat's Aunt Maggie
— He'll put a crimp in the cat

HE DISHED ME means
— He took my place
— He put me in a pan
— He jipped me
— He side-stepped me

SHOW YOUR SPOTS means
— You've got the chicken-pox
— How's your jack?
— Let's have your money
— Start something

THERE'S A YELLOW STREAK UP YOUR BACK means
— Your coat is in need of brushing
— You are not on the map
— You are a coward
— You haven't the nerve

A BIRD IN THE HAND IS WORTH TWO IN THE BUSH means
— Better be content with little than gamble for more
— An hour tomorrow is worth two today
— One crow does not make a winter
— He will hide his head under his wing, poor thing

HE'LL PUT OUT YOUR WINDOW LIGHTS means
— He will break your auto lights
— He intends to blind you

— He will biff you one in the eye
— He will give you the go-by

To Dope It Out means
— To think it over
— To ladle
— To get the hang of it
— To knock you dead

Half a Loaf is Better than None means
— Be satisfied with what you can get
— Bread is the staff of life
— Better poor with honor than rich with shame
— When she got there the cupboard was bare

See That Jane means
— I wish you'd hush
— There goes Jane, look!
— Look at the young lady
— Turn your lamps on the dame

Pull It Over means
— Throwing the monkey-dust
— Give it the once-over
— Close the lid
— Succeed in doing it

Lay Off That Stuff means
— Cut it out
— Place that material down
— Hold your horses
— Don't do it any more

To Walk a Chalk Line means
— To go straight
— To practise for the tight-rope
— To lead an honest life
— Don't fail me

Watch Your Step means
— Be careful
— Look where you are walking
— Keep your eyes peeled
— Some baby!

SAME-OPPOSITES

Look at these two words: yell . . . . holler Same Opposite

They mean the same, so a line has been drawn under Same.

Look at the next two words: long . . . . short Same Opposite

These two words do not mean the same; they mean just the opposite, so a line is drawn under Opposite.

Now look at all the other words on the page. If the words of a pair mean the same or nearly the same, draw a line under SAME. If they mean the opposite, draw a line under OPPOSITE. If you can't be sure, guess.

cold	hot	SAME	OPPOSITE
dragged-out	all-in	SAME	OPPOSITE
stall	jostler	SAME	OPPOSITE
razz	tease	SAME	OPPOSITE
hand-out	food	SAME	OPPOSITE
duck	disappear	SAME	OPPOSITE
dinky	nice	SAME	OPPOSITE
cop	bull	SAME	OPPOSITE
bonehead	bright	SAME	OPPOSITE
buck	plunk	SAME	OPPOSITE
crust	timidity	SAME	OPPOSITE
Murphy	potato	SAME	OPPOSITE
shyster	swell lawyer	SAME	OPPOSITE
suds	beer	SAME	OPPOSITE
bitter	sweet	SAME	OPPOSITE
take	accept	SAME	OPPOSITE
broad	man	SAME	OPPOSITE
fobber	fob-worker	SAME	OPPOSITE
poor prune	poor fish	SAME	OPPOSITE
hot dog	weenie	SAME	OPPOSITE
crust	nerve	SAME	OPPOSITE
rubberneck	deadhead	SAME	OPPOSITE
scratch	small change	SAME	OPPOSITE
slippery	easy mark	SAME	OPPOSITE
slang	watch-chain	SAME	OPPOSITE
flivver	tin-lizzie	SAME	OPPOSITE
tog	coat	SAME	OPPOSITE
stag	married	SAME	OPPOSITE
command	obey	SAME	OPPOSITE
flunk	succeed	SAME	OPPOSITE
poke	pocketbook	SAME	OPPOSITE
ancient	modern	SAME	OPPOSITE
wire	tool	SAME	OPPOSITE
low-down	classy	SAME	OPPOSITE
to lift	to pinch	SAME	OPPOSITE
hootch	moonshine	SAME	OPPOSITE
bestow	confer	SAME	OPPOSITE
rattler	trolley	SAME	OPPOSITE
dub	innocence	SAME	OPPOSITE
sky pilot	crook	SAME	OPPOSITE

whopper	lie	SAME	OPPOSITE
pickled	drunk	SAME	OPPOSITE
amiable	surly	SAME	OPPOSITE
cinch	snap	SAME	OPPOSITE
socking	hitting	SAME	OPPOSITE
pasteboards	cards	SAME	OPPOSITE
booster	hyster	SAME	OPPOSITE
rube	shark	SAME	OPPOSITE
lounge-lizard	cake-eater	SAME	OPPOSITE
prop-getter	stone-getter	SAME	OPPOSITE

CLASSIFICATIONS

Write the letter S under every word that is used in SMOKING.

Write the letter G under every word that is used in GAMBLING OR CARDS.

Write the letter D under every word that has something to do with DRINKING.

Write the letter T under every word connected with some form of THIEVING.

Write the letter F under every word that has to do with FIGHTING.

Write the letter A under every word that has to do with GETTING ARRESTED.

Write the letter W under every word that has to do with WOMEN.

Write the letter C under every word that has to do with CRAPS.

1. nicked, moll, four-flushing, jimmying, sand-bag, whopper, salt-creek.
2. deal, dead-head, snitched, booze, skirt, Little Joe, moll-buzzer.
3. slacker, nicotine, pulled, kick, scalaway, dive, rough-house, flivver.
4. weed, cinch, biff, snoozers, petticoat, fag, boister, rye, beak.
5. gorilla, pickled, coco, wollop, bad fall, bull, white line, hock.

6. Paddy's cabbage loaf, pennyweighter, pinched, broad, Come Phoebe.
7. Hen Medics, boot-legger, damper-getters, flush, tin-horn.
8. burglar, slugger, three-fingered, shimmy-shaker, swop, raisin.
9. jelly-bean, wren, stewed, punch, hoister, slugged, lamaster.
10. salve, flounces, pulled, spieler, mob, carousing, gopher-man.

11. magistrate, mugged, settled, screw, to be loaded, spook, peach, Come 7.
12. African Golf, stalling, Black Maria, fixer, female, black-jack.
13. porch-climber, toper, roll 'em bones, full house, rumbler, bracelets.
14. lammed, jay-walker, snitched, flossy, jazz, irrigate, date, lifer.

15. bammed, Fatima, hock, all blue, lambaste, sock soused, Sleepy-Hollow.

16. skirt, track No. 13, G. & W., brainstorm, trusty, jag, rappers, timer.
17. meerschaum, stagger, Baby needs shoes, camisole, trumps, injure.
18. coffin-nail, whiskey, phizog, chicken, anchor, dice.

OLD AND NEW METHODS OF TEACHING PRIMARY READING

EDITH M. PEYTON

Assistant County Superintendent, Athens, Ohio,

and

JAMES P. PORTER

Ohio University

The opportunity for this study was afforded by the fact that in Athens County two distinctly different methods of teaching primary reading, the one recent and scientifically grounded, the other older and not based on recent research, were being employed.

Any method which is today accepted as psychologically correct for teaching beginners to read should emphasize content rather than form; should call for thought as well as the more purely mechanical processes and should appeal to the child's imagination while it satisfies his natural craving for a story. That the aims of primary reading may be accomplished and efficient reading habits instilled, it is very important to have books which are attractive and which provide scientific and well balanced teaching methods carefully worked out in connection with real children's stories of high literary quality. Even the illustrations of modern texts have a very definite bearing on their usefulness.

The new methods, one of which was in its first year of use in Athens County, meet these demands very adequately.[1] Their readers, charts, flash cards and manual for teachers were

[1] The new method referred to is the Story Hour, published by the American Book Company.

published within the past few years in order to provide material in keeping with the results of the latest investigation in the field of teaching reading.

The older methods published earlier are more purely formal with the emphasis largely on phonics. The content is such as to be relatively uninteresting to the child mind, being made up of many short sentences composed to afford exercise of the phonetic combinations presented. The illustrations are in black and white, and costumes used are out of date. Word phrases which the child should learn to see and grasp as single units are broken up to meet the demands of the mechanical bookmaking.

The problem, then, was to ascertain and prove by means of standardized tests, whether children taught in a modern way, with attractive books, providing interesting content, really made more efficient readers than those taught in the older, formal manner with the emphasis practically all on the mechanical side.

For the purposes of this study, three schools of each type were selected. Those using the older methods will be designated as Schools A, B and C; those using the newer, as X, Y, Z. First grades chiefly were studied, but one third grade comparative study was included.

It was planned to give, at the outset, intelligence tests in all the schools studied, in order that the ability of the children taking the reading tests would be known. In this way it was felt the reading ability developed in the several schools, as shown by the results of the tests, could be more intelligently interpreted when considered in connection with the mental ability of these same schools.

The tests for mental ability—Detroit First Grade Intelligence and Pressey Primary Classification, were given early in January and reading tests, Haggerty Sigma I and Stanford Primary, Form A, were given the last of April, near the close of the school year. The tests except for the third grade, were all given by the same person. In every case, if the conditions under which they were given could not be identical, an effort

was made to see that schools A, B and C had the advantage. The teachers in schools X, Y and Z were handling their method for the first time, while the teachers in schools A, B and C had all had at least one year's experience in the method they were using.

It was realized at the outset that there were many things of value which could not be measured and would therefore leave this report tentative and incomplete. Such significant factors as the teacher's personality, the enthusiasm of the children, the correctness of the habits they form, their rate of silent and, in most cases, oral reading—none of these are dealt with in this study. We have attempted here a measure only of the native ability of the children and of their ability to comprehend what they read.

It was first planned to contrast schools A and X because the teachers had much the same experience and training. Both were rated as rather above the average. Schools B and Y were to be contrasted because their conditions were very similar, the teachers had about the same experience and training, each had one grade and both schools were in mining communities. Schools C and Z were two-room schools, each teacher having four grades in her charge. They were unlike in that C was a mining district while Z was agricultural.

Because the teacher is so significant a part of the school and determines so largely what may be accomplished with any method, data concerning the experience, amount of training and success of the teachers in the schools being studied are given in Table 1. Detailed study of the facts about the teachers shows that favorable and unfavorable conditions were more evenly balanced than studied prearrangement often produces.

Table 2 shows the true median group intelligence quotients for the different schools as established by the Pressey and Detroit tests. The Detroit test, when used as directed by its authors, classifies children into fast, average and slow moving groups. A GIQ (group intelligence quotient) for this test was obtained by finding a true median from the scores of 273 children tested in the city of Athens and Athens County, and then

TABLE 1

Experience, training and success of teachers

SCHOOLS	YEARS EXPERIENCE	HOURS TRAINING	RATING BY SUPERINTENDENT
A	12	32.5	Above average
X	22	36	Above average
B	1	66	Average
Y	1	89	Above average
C	2	47	Below average
Z	1	48.5	Average

dividing each score by the median thus found—36.6. The resulting GIQ's are higher than those from the Pressey test, which can be accounted for because the median was established from local cases only and because while the Detroit is a test designed to be given in September to children just entering school, it was given here in January with the result that scores were a little higher, no doubt, than if the test had been given in September. Except for comparison with Pressey medians, Table 2, which shows that in most cases they confirm the Pressey results, no further use will be made of the Detroit results in this report.

TABLE 2

Median GIQ

SCHOOL	NUMBER OF CASES	PRESSEY	DETROIT
A	35	94	101.5
X	11	90.9	102.8
B	34	90.3	103.5
Y	21	91.5	105.5
C	5	79.5	79.5
Z	6	109.5	119.5

It will be noted that GIQ's in general are low, indicating very clearly that the level of native mental ability with which the teachers worked was probably below standard. The scores obtained from the reading tests could therefore not be high.

In considering the results of the reading tests, it seemed most significant to find both the median and the average scores, of which records are made in Table 3.

TABLE 3

Median and average scores of reading tests

SCHOOL	HAGGERTY READING			STANDFORD ACHIEVEMENT A		
	Number of Cases	App. Mdn.	Average	Number of Cases	App. Mdn.	Average
A	34	7	7	28	5	6.3
X	10	6.5	6	7	4	4.1
B	31	3	3.5	24	3	3.5
Y	18	9	10.16	18	5	6.6
C	5	2	2.8	Test not given		
Z	5	9	11.6	7	4	9.5

To place the results of measurements of reading ability on a more readily comparable basis, the per cent of cases in each school with GIQ's below 100 was ascertained from Pressey Primary; also the per cent of cases that ranked, according to the Haggerty standard, at standard and above first grade reading ability. If any group of pupils has a large per cent testing below 100, this of course means a more difficult task for teacher and method to bring such up to standard or above in reading or other achievement. Therefore we have computed the per cent below 100 in mental ability in order to facilitate comparison. The Standard Achievement reading medians and averages as will be learned from Table 3 are directly comparable with the same values for the Haggerty tests. They demonstrate much the same direction and amount of differences. Because of the

absence of standards for first grade children and of results for school C in which the Stanford tests were not given, later detailed comparisons will be made in terms only of the Haggerty reading tests.

TABLE 4

Per cent at and above standard in reading and per cent below in intelligence

SCHOOL	PER CENT AT AND ABOVE HAGGERTY STANDARD	PER CENT BELOW 100 GIQ PRESSEY
A	66	85
X	40	75
B	15	86
Y	77	76
C	20	100
Z	83	33

Table 4 shows the per cent in each school ranking above the standard established by Haggerty as normal first grade reading ability.

School A, with 85 per cent below 100 in GIQ's, using the old method, succeeded in teaching reading so that 66 per cent rank, according to Haggerty, at standard and above for first grade ability. School X with 75 per cent below 100 in GIQ's, produced 40 per cent at standard and above in reading ability. This case taken alone, shows the superiority of the old method.

School Y, with 10 per cent higher intelligence than School B, using the modern method, produced 62 per cent more readers measuring at standard and above. That is, School Y shows 77 per cent at standard and above in reading ability, while School B shows 15. It will be remembered that both these schools are in mining communities, and that there is a difference of only 10 in percentage of GIQ's below 100.

While between schools C and Z there is a wide difference in mental ability—100 per cent below 100 in GIQ's in school C and 33 per cent in school Z, there is an even greater difference in reading ability. School C had 20 per cent at standard and

above as contrasted with 83 per cent at standard and above for school Z.

In the last two pairs of cases the evidence is strongly in favor of the modern method.

Schools A and Y also should be compared with each other, as A and X with each other, because in these two schools we we find rated equally high teachers in each method and each is located in more comparable communities industrially.

Careful comparative study of Table 4 shows school Y to excel school A in both reading achievement and mental ability, but somewhat more in reading ability. The same results are obtained if schools B and X are compared, though in this case there are not the same facts about teachers and communities for making the comparison.

If we average the three new method schools in per cent at standard and above first grade reading ability, we find them 66 per cent standard and above to contrast with 33 per cent for the old method schools. Treating the per cent with intelligence quotients below 100 in the same way, we find the new method schools with an average of 61 per cent below to contrast with 90 per cent for old method schools. This means that new method schools, with an advantage of 29 per cent in ability, show results 33 per cent better in reading achievement.

STANDARD	NEW METHOD SCHOOLS	OLD METHOD SCHOOLS	NEW METHOD ADVANTAGE
Average per cent standard and above in reading ability.....	66	33	33
Average per cent below 100 in IQ.....	61	90	29

No study like the present one is complete which does not include finally the investigation and treatment of individual children. The next step, therefore, was to select all those children having identical Pressey GIQ's. Those from the new method schools were grouped together and the average Hag-

gerty reading score of 9.6 obtained. Children with identical GIQ's, but taught in old method schools, have an average Haggerty reading score of 4.8. To be sure, the number in each group is small,13 in the new method and 18 in the old. This method of dealing with individuals of the same mental ability seems to us so much more fair and convincing that we are disposed to give considerable emphasis to it and results obtained by it. Our findings to date would indicate that new methods of teaching reading are about twice as efficient as the old.

A third grade in Athens County which had used the old reading method since entering school was able to secure the new method text and some materials for use during the second semester of the school year 1924-25. The teacher and supervisor, in conference before the new books arrived, agreed to let the fact that the class was beginning a new book motivate an active campaign to improve the quality of their reading. The children were told of the plan and urged to put forth effort to see how much improvement could be made in a given time.

A record of the mental ability of the class had been secured about a year earlier by the use of the Pressey Primary group test for intelligence.

A standardized test was used to determine whether there was a growth in reading ability. The Stanford Achievement Primary, Form A, was given at the beginning of the period of especial endeavor and Form B at its close, six weeks later.[2]

The same tests, Pressey Primary Classification and Stanford Primary A and B, were given on approximately corresponding dates to another third grade where no particular campaign for improvement was being instituted. This group was to serve as a check or control and, representing the usual traditional procedure, help to show whether arousing in the children of the experimental group a conscious desire for improvement and stimulating them with new material and carefully planned drills brought about increased ability to read. Hereafter in

[2] We are indebted to Dorothy J. Slutz, graduate student, for giving, scoring and working up the results of these Stanford Achievement tests.

this paper the third grade in which the definite conscious attempt was being made to improve reading will be called the experimental, the other, the control group.

In the experimental group each child's silent reading rate was ascertained before the new book was begun. Tests to determine rate were repeated at intervals of two weeks. Informal tests for comprehension were made frequently.

The eye span of these children was at first very limited. They seemed actually unable to see more than a word at a time. So an effort was made to widen the eye span by using flash cards with word groups. Games using word group flash cards were employed. As a part of the preparation for each lesson in the new book, phrases and word groups were drilled upon, usually in the spirit of a game and after their meaning had been learned. To aid in the development of the weakest cases, a series of phrase books was prepared. Each book consisted of ten phrases. The books were graded in difficulty from that of the Primer to the Third Reader. When a child could make a perfect score in one book, as the phrases were flashed for him, he was permitted to try the next.

The teacher and the supervisor held conferences each week, studying the progress being made by the group and individuals, diagnosing difficulties and planning remedial measures.

The children in the control group had also used the old reading method from the time they started to school. Somewhat prior to the time of our comparative study they were using new method readers as supplementary material. The experimental school had had no such supplementary material before this period of experimentation began.

In the control school no effort was made to arouse in the children a conscious desire to "beat their own record". The teacher was left largely to her own resourcefulness in planning and motivating her work.

The results of the tests of mental ability indicate that the control school stands somewhat higher. The approximate median of the Pressey group intelligence quotients for each third grade is: experimental, 85; control, 91.

The results of the Stanford Achievement reading test are in agreement with those of the intelligence tests, the experimental grade having lower scores than the control grade, the approximate median scores of the former being 19 in Form A and six weeks later in Form B, 36; in the latter, 35 in Form A and in Form B, 40.

The experimental grade contained many retarded children. Their chronological ages ranged roughly from eight to fourteen years. This made quite a difference in computing the Educational Quotient from these scores. The Educational Quotient (EQ) is obtained by dividing educational age as found by test by the chronological age, and multiplying by 100 or in exactly the same way as the intelligence quotient is obtained. The median EQ's of the two schools is as follows:

	Third Grades	*Form A*	*Form B*
Median EQ's	Experimental	79	82
	Control	90	91

The control grade gained 1, and the experimental, 3 points in median educational quotient. It is very interesting to note the number of cases with higher scores and educational quotients (EQ's) on the second test. In the experimental group, 80 per cent of the children made higher scores on the second reading test than on the first. In the control group 43 per cent of the scores were higher on the second test. The EQ's for the two grades show that in the experimental 85 and in the control grade 47 per cent gained on the second test. The evidence of gain in ability to read as measured by the Stanford test is clearly greater in the experimental group. This gain, it should be emphasized, was made by less able pupils by means of definite motivation, drill, a new method psychologically determined, by a teacher of no more than average ability, but willing to cooperate fully with a trained supervisor.

The rate of silent reading was not measured in the control grade. This was done at frequent intervals in the experimental grade. While, therefore, no comparative results can be given, it is significant to find that as in other elements of reading

ability, this experimental third grade at the first ranked well below standard. Only 25 per cent were at or above standard for rate of silent reading as given by Courtis for the third grade. This standard is 135 words per minute. At the close of the six-week experimental period, 54 per cent were at or above this same standard. Inasmuch as this gain in rate agrees in both direction and amount with the other measures of reading ability we may conclude again that our experiment was successful and worth while.

Summary of Results and Conclusions

1. Except in one instance every comparison of two first grades, the reading achievement of the children who had learned by the newer and more scientific psychological method was superior. There were as a matter of fact strong unfavorable conditions in this one instance which could well cancel the greater efficiency of the newer method. Though the native mental ability of the grades using the newer methods happened to be higher, the reading achievement in those learning by the newer methods was higher yet.

2. Individual children having the same mental ability as determined by tests when learning to read by the newer method show about twice the reading achievement of those learning by the older methods.

3. A third grade selected for definite experimental procedure consciously set to use the new method with all its appeals to favorable mental and physical processes involved in learning to read made a considerably greater gain than a control third grade which, except for incidental supplementary reading work, made no use of the newer methods. This gain was made in the face of a lower average mental ability in the experimental grade. There were also other adverse conditions. A six-week period is not too short to demonstrate what a new method supplemented with definitely working motivation of teacher and pupil can accomplish.

4. The silent reading rate was so increased that more than

twice as many pupils reached or exceeded the standard rate of 135 words per minute.

5. This study would have been most difficult, if not impossible, without the work of the supervisor of primary reading. The training of teachers in newer and better methods, the giving of tests, scoring and working up results, advising and inspiring teachers and pupils to higher achievement, all this and more, can come most readily for rural, as it has for city children only through scientific supervision.

6. In several earlier attempts to investigate the efficiency of supervision and the newer methods of learning to read, no definite tests of native mental ability of the children were made. Starch summarizes admirably (Ed. Psych., Chap. 16) the methods and results of the studies of the effectiveness of special methods of teaching reading. He rightly concludes, as we do in this study, that there are many complex factors involved and that no final conclusions can be drawn.

The Indiana experiment of gains through supervision of rural schools in two counties with two other counties used as controls shows clearly improvement in achievement in reading and other subjects.

In none of these investigations just mentioned was there an attempt made to determine the mental level of the children on whom various methods and supervision were used.

Gray in his excellent and thorough Summary of Investigations Relating to Reading (University of Chicago, 1925), reports results of studies of the relation of intelligence to reading achievement. Such workers as True, Madsen, Merrill and Hewes with grade and special class pupils and Webb with college students find a greater or less degree of positive agreement between general intelligence and reading achievement. Theisen, Oglesby, Lowell, Dickson with younger pupils and Sangren and Patrick with older pupils and college students demonstrate that progress in reading depends to a considerable extent on the level of intelligence of the learner and also, what is more germane to our work, that further experiments are needed to determine the most appropriate types

of instruction for pupils of various levels of intelligence and at various levels of school progress.

If our own first results of measuring efficiency of newer reading methods as against older ones seem to the reader unconvincing, may we hope that he can approve of a method which first requires that the mental ability of children be ascertained before undertaking to determine the advantages of various methods of training. In so concluding, it is by no means necessary to assume that the intelligence tests used by us were perfect measures of the mental ability of our children.

THE PERCEPTIBILITY OF AUTOMOBILE LICENSE NUMBERS

JAMES D. WEINLAND

Lehigh University

In the last few years the numbers of automobiles have been increasing so rapidly that in several states they have passed the million mark. Each car carries a license plate with a number and since the cars have become so very numerous these license numbers have become quite cumbersome and difficult to read.

It seems that psychology ought to be able to find a way to simplify and clarify this situation. Theoretical studies in perception have been carried on for many years, and especially in regard to reading has valuable knowledge been gathered. Some of these known facts might now be applied to the automobile license problem.

The immediate introduction of the writer to this question was brought about by a study at Columbia University for his Master's thesis on "The effect of grouping upon the perception of digits." He found in this work, much to his surprise, that the 3–3 grouping of a series of six digits was inefficient and tended to error. This finding immediately suggested automobile license plates where the 3–3 grouping is often seen. As the Columbia work was done with a small tachistoscope, the digits used being typewritten and exposed at a distance of 3 or 4 feet, the conclusions could not be applied to automobile license numbers without further experimentation.

The present investigation was started to test the effect of grouping with actual automobile license numbers, but developed into a study of the effect of using letters on the license plates with the numbers.

The apparatus used in the experiment was a special tachisto-

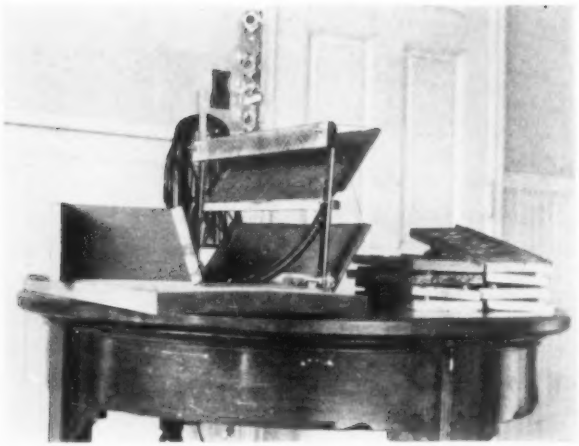


FIG. 1. REAR VIEW OF TACHISTOSCOPE, SHOWING BOTH DOORS PARTIALLY LOWERED

FIG. 2. FRONT VIEW, DURING THE EXPOSURE

scope designed by the writer. It consisted of two doors hinged horizontally and mounted the one above the other. In operation the bottom door would drop, exposing the numbers, and pull the upper door, by a cord, after it, ending the exposure (see the pictures). The exposure time used in this experiment was approximately one-fifth of a second. It was not measured accurately, but was kept constant throughout the whole experiment so that for comparative purposes the results are entirely reliable.

The particular digits shown in each exposure were selected by chance. Digits 1, 2, 3, 4, 5, 6, 7, 8, 9 and 0 were written on small cards and shaken up in a box. The requisite number of cards for the exposure were then drawn out and the digits written on them taken in their order.

The automobile license numbers were furnished by the Pennsylvania State Bureau which had charge of their manufacture and were identical with the numbers used on the Pennsylvania automobile license plates. The numbers were sent separately, however, so that the order of presenting them could be changed.

The letters used in the experiment were purchased in a hardware store. They were a trifle smaller than the numbers but it was judged that the difference in size was too small to materially affect the experiment. The letters and background were painted the same colors as were the digits and their background (yellow background, blue letters and digits).

Before every experiment a series of exposures was prepared. Either the digits, or digits and letters were mounted by a pre-arranged scheme upon boards 6 inches wide which in length just fitted the tachistoscope. During the experiment one of these boards was slipped in place, the ready signal given, the exposure, and then the board was taken out and another one with a different grouping put in its place. The groupings were shown twice, in chance order, during each experiment.

The subjects were a group of college men, largely of the Junior class who were taking a course in Applied Psychology. Their coöperation was good. The regular class situation and

laboratory periods were used. The experiment was carried on in a large room the men being from 15 to 25 feet away from the tachistoscope. The men kept the same seats throughout the whole experiment to make the results accurately comparable.

The actual experimentation was done by two senior men Borda and Gorgas. These men also did the major work in compiling the results.

The first half of the experiment dealt with the grouping of the digits and will not be quoted here as the differences between the groupings under the conditions in which the experiment was made, were not great enough to be very significant. Six digits as a single group appeared the best, but it seemed that results of greater reliability and usefulness ought to be available.

It was at this point that the idea developed of using letters. Previous tachistoscopal work had indicated that letters combined in meaningful syllables or words were much more perceptible than digits. Why then, should letters not be given a numerical significance and used in numbering automobiles?

Below the results are given of a test of this hypothesis. A brief preliminary experiment showed that the accuracy of report with digits up to five in number was as high as 90 per cent. Letters were then introduced in nonsense syllables and small words. The combinations used were as follows.

Letters	Numbers	
0	6	
0	5	
3	3	
3	3	(numbers first)
3	4	
4	2	
5	2	
1 digit—1 letter—4 digits		(New York Plan)

More combinations could well be used to make the experiment quite complete, and probably will be next year. The results obtained were sufficiently decisive, however, to justify conclusions which seem to the writer of considerable practical importance.

Some sample exposures are given below with the percentage of correct responses, as found in the experiment.

"Correct" means the whole number was recorded without error.

"Exposures" means the number of combinations used (50684; one combination) times twice exposed, times the number of subjects who recorded. (The two exposures were separated by exposures of other combinations to eliminate memory responses.)

6 digits. 56 per cent correct. 200 exposures.
Samples: 473926
081576
359204
726358

5 digits. 90 per cent correct. 680 exposures.
Samples: 50684
48692
62408
13758

3 letters—3 numbers. 70 per cent correct. 400 exposures.
Samples: DIM125
MAY461
TIK958
BEL291

3 numbers—3 letters. 76 per cent correct. 288 exposures.
Samples: 543BAD
193PET
195KOB
462MUT

3 letters—4 numbers. 45 per cent correct. 720 exposures.
Samples: PUD7145
DOT6309
SOL3618
BUD1603

4 letters—2 numbers. 85 per cent correct. 266 exposures.
Samples: BALD19
PALE46
TIDE35
BAKO19

5 letters—2 numbers. 75 per cent correct. 330 exposures.
Samples: BATUK96
DOUBT82
MOLDY73
LIMED56

In the above exposures there was a small space allowed between the letters and digits, but it was not a full space and cannot be shown with a typewriter.

1 digit—1 letter—4 digits. 35 per cent correct. 260 exposures.

Samples: 2B 70 64
41 95 18
9A 37 54
7M 39 06

Letters	Numbers	Per cent correct	
0	6	56	
0	5	90	
3	3	70	
3	3	76	(reverse)
3	4	45	
4	2	85	
5	2	75	
1 digit—1 letter—4 digits		35	

This experiment will be carried on further to get more refined results, but the figures quoted here show plainly the high perceptibility of letter groupings.

It merely remains to evolve some plan of giving numbers a numerical significance to show their utility for automobile license plates. One simple plan using the experimental data given above might be as follows.

99,999	5 digits
519,948	1 letter and 4 digits
419,580	2 letters and 3 digits
4,305,590	3 letters and 3 digits
4,365,900	4 letters and 2 digits (not exhaustive)
10,914,750	5 letters and 2 digits (not exhaustive)
20,625,757	total

99999 cars numbered with digits alone.

1 letter and 4 digits used to number 26 (letters) times 9999 times 2 (the letter used both before and after the digits.) This would number 519,948 cars.

2 letters and 3 digits. 21 letters times 5 vowels times 2 (AB-BA) times 999 times 2 (letters used before and after the digits). This would number 419,580 cars.

3 letters and 3 digits. 21 consonants times 21 consonants times 5 vowels times 999 times 2 (letters used both before and after the digits). This would number 4,305,590 cars.

4 letters and 2 digits. 21 consonants times 21 consonants times 5 vowels times 5 vowels (abet) times 2 (bate) times 99 times 2 (letters used before and after the digits. This would number 4,365,900 cars but is not exhaustive for the four letter and 2 digit combination, as the last consonant could be doubled (bett) and many small words were not counted, beam for instance.

5 letters and 2 digits. 5 vowels times 21 consonants times 5 vowels times 21 consonants times 5 vowels times 99 (ababa) times 2 (letters used before and after digits). This would number 10,914,750 cars and is by no means exhaustive for the combination.

A scheme as the one suggested could be completely worked out to number the cars of any state or nation and the system would easily take care of more than twenty million cars. A 25 per cent increase of efficiency in perceptibility over 6 digit license numbers should be found with all the new numbers and a much greater than 25 per cent improvement with many of them. The scheme would be applicable also to numbering freight cars, or any series of things where quick and accurate perceptibility of the number is desirable.

REFERENCES

CATTELL, J. M.: Ueber die Trägheit der Netzhaut und des Sehcentrums. Phil. Stud., 3, 1886, 94ff.

DEARBORN, W. F.: Perception and Reading. Psy. Res. of J. M. Cattell, 1914, 37.

DODGE, R.: Psychology of Reading. Psy. Rev., 8, 1901, 56. Study of Visual Fixation. Psy. Rev. Mon. Sup., 1907, No. 35.

HUEY, E. B.: Psy. and Ped. of Reading, 1908.

WEINLAND, J. D.: The effect of grouping on the perception of digits. Amer. Jour. Psy., April, 1924, vol. 35, 222ff.

WHIPPLE, G. M.: Simpler Processes, p. 266ff.

NOTES AND NEWS

THE AMERICAN HISTORICAL ASSOCIATION

The American Historical Association is now for the first time making a general appeal to the public for coöperation in the raising of an adequate endowment. The Association was founded in 1884 by such men as Andrew D. White, Francis A. Walker, and Herbert B. Adams. The presidency of the Association has been held by two Presidents of the United States—Theodore Roosevelt and Woodrow Wilson. The Association is not an exclusively academic institution. It is the chief coördinating agency of the historical fraternity throughout the country. It has stimulated and guided the teachers of history in elementary and secondary schools. It is now, in view of changed conditions, planning a fresh study of these educational problems.

The Association undertakes to place at the disposal of the Government the resources of historical scholarship. It should be noted that the prizes administered by the Association are given for studies in European as well as American history.

Such work as the Association has already done has been made possible only through an immense amount of unpaid service. What the Association now asks is an increase in endowment from $50,000 to $1,000,000 with the expectation that the additional income thus provided will be used, not only to secure more certain and adequate support for work already undertaken, some of which has been seriously curtailed or delayed by the lack of such support; but also to make possible certain new forms of service.

The time has now come, however, when the Association ought to assume a more positive leadership. We have not had in this country enough systematic effort to direct the attention of investigators toward special problems which need to be solved, whether from the point of view of the historians themselves or with proper regard for workers in the related social sciences, such as economics, political science, and sociology.

Now that the United States has become the most influential of the World powers, there is urgent need for more thorough and dispassionate studies of the history of American international relations (including contacts with Europe, Latin-America, and the Far East). The appeal to history is often made by partisans, but nearly always such appeals involve more or less distortion of the facts. There are few subjects more in need of disinterested research.

A topic much obscured by prejudice and emotional or selfish propaganda is that of immigration and racial elements. Recent legislation brings up questions in which the help of the critical historian as well as of the economist and the sociologist is needed.

Although the Association can always draw extensively upon its members for unpaid professional services, the desirability of having some one devote the major part of his time to directing and correlating its activities is very evident. Funds are needed for the salary of such an official, and also for the rent of offices, or perhaps the purchase and upkeep of a house in Washington to serve as headquarters for the Association and to house properly its accumulating records.

Any one interested should write to The American Historial Association, 1140 Woodward Building, Washington, D. C.

The Southern Society for Philosophy and Psychology

The Twenty-First Annual Meeting of the Southern Society for Philosophy and Psychology met at the University of Kentucky, April 2 and 3, 1926, Dr. J. B. Miner, President, presiding. The program of the meeting was so varied and interesting as to provoke vigorous discussion from the first to the last paper. President Miner's address on The Significance for Psychology of Qualitative Differences was most thoroughly and thoughtfully prepared. It was preceded and followed by discussions of Behaviorism and Systematic Psychology by Drs. Max Meyer, A. P. Weiss, I. Madison Bentley, Knight Dunlap, Herbert Sanborn, Seba Eldridge, Joseph Peterson and other guests and members of the Southern Society. Altogether the proceedings of this Society, at least for this year, will compare favorably with those of the American Association.

At the business session the following resolutions were adopted:

Whereas a study of the teaching of psychology in colleges, state teachers' colleges and normal schools in the South reveals that psychology is taught in a number of institutions by teachers whose training has been inadequate; and whereas the equipment of laboratories, technical periodicals, etc., necessary for proper training in psychology has also been found to be inadequate (conditions which probably do not obtain in the South alone); therefore, be it resolved that:

1. The Southern Society for Philosophy and Psychology hereby appoints a special committee with power to carry out these resolutions, consisting of Professor Joseph Peterson, of the George Peabody College for Teachers, as Chairman, and Professor Buford Johnson, of the Johns Hopkins University; Professor J. B. Miner, of the University of Kentucky; Professor T. K. Sisk, of the State Normal School, Livingston, Ala.; and Professor W. J. Woodson, of the State Teachers College, San Marcos, Texas.

2. That the Society express its disapproval of the teaching of the first year's courses in psychology, educational psychology, and other branches of psychology by anyone who has not had the training represented by at least a master's degree with a major in psychology; and also the teaching of advanced courses in these subjects beyond the first year's courses by anyone who has not had at least the training equivalent to a doctor's degree with a major in psychology.

3. That the Society asks all heads of departments in which these subjects are taught, whenever the question of transfer of credit from other institutions to theirs is raised, to disapprove the transfer of any credit in these subjects not taught by persons qualified as indicated above.

4. The Society asks the above Committee to coöperate with the various institutions and the associations of colleges in every way possible in raising their standards in accordance with the above resolutions.

5. That the Society authorizes the Council to appropriate such sums as are necessary for carrying out the intentions of the above resolutions.

The Brown-Woody Civics Test

The World Book Company has recently published the Brown-Woody Civics Test, designed for measuring the achievement of students in high school civics or in community civics of seventh, eighth, or ninth school years. With the importance attached to citizenship training in schools, it is believed that an objective measurement of the results of civics teaching is highly desirable. The items of each of the three parts have been chosen with great care. Standardization has been experimentally and thoroughly done. This test should prove useful for classifying, measuring class and pupil progress, diagnosis of difficulties of class or student, testing methods of instruction, and the motivating of learning and stimulating of teacher and pupil.

Courses on Home Education and Parenthood and the Pre-School Child

Dr. Garry Cleveland Myers was on leave from the Junior Teachers College of the Cleveland School of Education during the past semester. He offered two courses in Home Education and Parenthood at the School of Applied Social Sciences, Western Reserve University to 136 regularly registered students, of whom 102 were fathers and mothers. Each course carried two semester hours credit and met in two sections. "Education of the Child of Pre-school Age" had 59 students and "Home Education of the Child from Six to Twelve" had 77. The course on the pre-school child is now in its third semester, being offered first in February 1925. The total registration in this course from that date has been 179 students.

The Sachs Prize Awarded

The Committee on the award of the Julius and Rosa Sachs prize of $1000, established at Teachers College, Columbia University, for the purpose of promoting the progress of secondary education in the United States has announced as the winner of the prize for 1926 Dr. Edward Augustus Fitzpatrick, Dean of the Graduate School, and Professor of Education, Marquette University, Milwaukee, Wisconsin. The subject of the essay is: "The Promotion of Scholarship in the Teachers of the Secondary Schools of the United States."

The second annual competition for the Sachs' prize of $1000 has already been announced. The topic of the essay or treatise for this second prize is: "The Aims and Methods of Science Teaching in the Successive States of the Secondary School, and the Intellectual Equipment of the Teacher That Will Enable Him to Put These Aims into Practice." All manuscripts must be in the hands of the Dean of Teachers College on or before December 1, 1926. The rules governing the competition for the Sachs' prize may be secured from the Secretary of Teachers College, 525 West 120th Street, New York City.

Army Alpha Revision

Dean R. Brimhall, Secretary of the Psychological Corporation, 3939 Grand Central Terminal, New York City, has announced a revision of Army Alpha by Dr. Bregman with the coöperation of Dr. Cattell. It is a combination of the five army forms, made to be of average difficulty and to give average scores. The best questions for general use have been selected by order of merit methods, thus excluding those addressed especially to men recruits at a particular time.

The test was prepared especially to be given by correspondence to those who are unable to consult a competent psychologist. Full instructions have been prepared so that any one can arrange to have the test made on himself. A charge of five dollars is made for the test. Material, scoring and a report of the rating made in the test will be mailed to the person concerned.

While the revision was taken up to fill the need for a test that might be given by correspondence, it is apparently the best form of Army Alpha for a group.

One hundred copies with mimeographed instructions can be supplied for five dollars to psychologists associated with the Psychological Corporation. It will not be sold to others.

Professor Norman Fenton, State Teachers College, Tempe, Ariz., has been appointed to an Associate Professorship in Psychology at Ohio University.

Stuart M. Stoke has accepted an appointment to an Assistant Professorship in Psychology in the University of Buffalo.

Dr. Dean A. Worcester has accepted appointment to an Associate Professorship of Psychology at Ohio University, Athens, Ohio. Dr. Worcester was formerly Professor of Psychology and Director of the Bureau of Educational Research in the State Teachers College, Emporia, Kansas. During the present summer he is Assistant Professor of Psychology, Ohio State University.

Dr. George E. Carrothers, Professor of Education, Ohio University, has been appointed Dean of Rollins College.

BOOK REVIEWS

RICHARD B. FRANKEN. *The Attention Value of Newspaper Advertisements.* New York, The Association of National Advertisers, 1925. Pp. 55.

Franken reports a summary of five separate investigations regarding the attention value of newspaper advertisements. Two of the studies were made by Franken at different times, one by Strong, a fourth by Roberts, and a fifth by Hough. In the appendix is presented a bibliography of thirty-four titles on the subject.

The reviewer is naturally pleased to find that the conclusions of his original study made in 1914 are well supported by these four subsequent investigations. In fact the most striking fact in Franken's report is that investigations based on from 25 to 102 persons should agree so well with each other.

A few of the many conclusions may be mentioned. "Although it is not likely that every page in the newspaper is read or even looked at by every person tested, nor that any advertisement, no matter what its size, can hope to obtain 100 per cent attention, nevertheless the results of the tests reported show that even so small a group as 100 typical readers is taken as a whole, some advertisements on every page are seen and recognized by some one person. The poorest page receives an attention average of 3 per cent, while the pages throughout the entire paper, excepting pages 2 and 3, receive the potential average attention of about 5 per cent." But some advertisements were not recognized by a single person tested. Thus does science cool the ardor of the sales promotion manager's talk of 100 per cent efficiency and call attention to the fact that some advertisements are woefully weak.

Another conclusion that is not welcomed by sellers of advertising space is that attention value increases approximately as the square root of the area and not proportionately to the area. This fact, first stated by Hollingworth, seems to be well established now. The effect of preferred position, of reader's reading habits, of position on the page, of the presence of reading matter, of size and shape of advertisements, and many other factors are considered.

One problem that is not discussed is that of the effect of size of the publication upon advertising efficiency. Years ago the reviewer published data showing that the memorability of an advertisement decreased as the number of advertisements seen at one time increased.

Today many advertisers are convinced that a page has not the advertising value it used to have, due to this factor. Here is an interesting psychological problem that needs further investigation.

No one interested in advertising can afford not to consider carefully the results so clearly presented by Franken, and in such compact form.

EDWARD K. STRONG, JR.
Stanford University.

EDWARD K. STRONG, JR. *Psychology of Selling and Advertising.* New York, McGraw-Hill Book Co., Inc. 468 pp. Illustrated.

On the whole, the book fortunately, in a field that already has so many books on psychology of advertising, stresses the principles of psychology applied to *selling*, although it also applies the same principles to advertising as well.

The purpose of Dr. Strong's book is "to supply a working formula in terms of which a seller may not only make each individual sale in an efficient manner, but may also tend steadily to convert buyers into customers." Since the fundamental objective of selling is to make a customer, he considers the psychological problems of selling in terms of permanent satisfaction, developing good will, and converting potential prospects into regular customers. Unless the seller can understand when certain changes have taken place in the minds of his prospects that cause them to have a new attitude toward the product, he cannot succeed in selling.

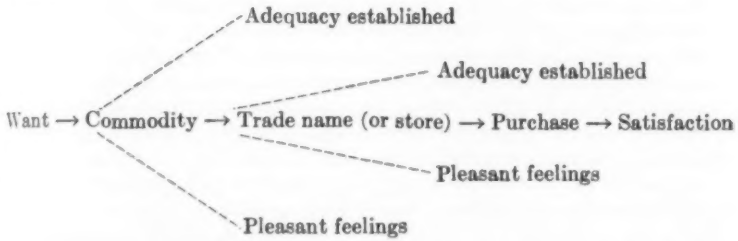
The book is written and arranged to meet the needs of three groups of people. Leaders in the advertising and selling fields can find all the information they need in Chapters 1 to 6, "General Survey." Sales "cubs" and students should begin by reading Part IV, "Strategy and Tactics of Selling." Business men, wishing to understand the basic factors involved in handling and influencing other people, will find their problem discussed in Part II, "How Man Satisfies His Wants," and Part III, "Principles of Selling and Advertising."

In commenting on changes in selling procedure that have been made in the past fifteen years, Dr. Strong stresses the growing recognition of the buyer's point of view. This change has come about so gradually that many salesmen and advertisers have not yet seen that if they want the buyer to be satisfied and feel good will, they must present their proposition from the buyer's point of view. A concrete instance is of a salesman whose product was a new fashioned ice-box with several patented features. Although the man's territory was in the South, he was not particularly successful until the sales-manager sent an ice-box to his home as a gift. When the salesman used the ice-box, he understood its good points and could make more sales because he had both factory knowledge and intimate contact with it. The salesman

had moved away from the trial-and-error method of selling, because he knew what his prospects wanted.

In Chapter 9 is found a discussion on the instruction regarding the commodity that companies give to their salesmen. All companies require some knowledge of the product. This varies from the formal instruction, typical of many companies, to the fourteen months intensive training in its own plant that the Westinghouse Electric and Manufacturing Company gives its prospective salesmen.

The salesman needs to understand that in buying anything, the purchaser's mental steps are "from want to commodity, to trade name, then purchase." When the buyer has used the commodity, he is either satisfied or dissatisfied. If he has found the commodity both adequate and pleasing, the buying formula can be expressed in a diagram, like this:


Adequacy established
Adequacy established
Want → Commodity → Trade name (or store) → Purchase → Satisfaction
Pleasant feelings
Pleasant feelings

In the next three chapters the buying formula is exemplified in five analyses of sales—actual interviews between the seller and the buyer. These are sales in the life insurance field, in retail groceries, and in the home. In each instance, the salesmen find the buying formula of value in understanding the mental process of buyer. They are salesmen who have had the four kinds of instruction that all salesmen should have. They are, first, knowledge of the commodity and the policies of his company; second, a knowledge of selling methods; third, ability to carry out the methods; fourth, understanding of his prospects.

Chapter 5 points out that "the solution to each sales problem is implied in the buying formula;" that is, that the same tactics of salesmanship can be used for various prospective buyers, but that emphasis must be shifted. The chapter on "where to put the emphasis" reproduces twenty-five advertisements of well-known products and points out in what respects they do or do not fit the buying formula.

In Part II, "How Man Satisfies His Wants," is given an interesting, non-technical discussion of man's natural wants and his normal means of satisfaction.

Part III, "Principles of Selling and Advertising," discusses motiva-

tion, its theory, practise, and technique; also the rôle of memory and the use of appeals.

Part IV, "Strategy and Tactics of Selling," gives clear and comprehensive instruction to salesmen. This section of the book is not intended to take the place of a company sales manual with its discussion of the firm's peculiar selling problems. But it will help the salesman understand more fully what is back of the instructions in his company's manual. There are helpful chapters on prospecting, starting the interview, etc., with helpful suggestions for every step that the salesman has to take. Dr. Strong understands the salesman's viewpoint. In his preface he speaks of having gained a broader understanding of the problems of selling from the several hundred experienced salesmen who have been in his classes.

Since Dr. Strong's book is based upon new objectives of selling that have arisen almost within the past decade, it presents a different view from those of older books on selling and advertising. It has modified old concepts and introduced many new ones. "Psychology of Selling and Advertising" will undoubtedly receive a warm welcome in the business world, for it will be helpful both to corporation heads and to their salesmen, as well as students of sales and advertising psychology. It is written from the viewpoint of both the seller and the buyer, has human interest, is full of concrete examples of the application of psychological principles to sales problems, and is very readable.

RICHARD B. FRANKEN,
New York University.

G. V. HAMILTON. *An Introduction to Objective Psychopathology.* Mosby & Company, 1925. 354 pp.

This book consists of twelve chapters and is divided into two parts. Part one is composed of three chapters: I, Introduction; II, Two Hundred Nervous Cases; III, Summary of the Survey Findings. The description of the 200 nervous cases covers 170 pages, averaging therefore, somewhat less than a page to a case.

Part two is entitled "Principles of Objective Psychopathology," to which the remaining nine chapters are devoted. Chapter IV is another Introduction; chapter V, The Foundations of Psychopathology chapter VI, Neural Morphology, Neural Physiology, Endocrinology; VII, Comparative Studies of Reactions to Baffling Disadvantages; VIII, Habit Formation; IX, The Relation of Inhibition of Responsiveness to Indirect Responsiveness; X, Unsatisfied Major Cravings; XI, Reactions to Inferiority; XII, Sexual Behavior. There is a glossary of unfamiliar terms, a bibliography and an index.

In a foreword, Dr. Yerkes says of the book that it "will be rudely handled by many, neglected by others, greedily devoured and used as

a stepping stone to fuller knowledge and the development of methods of observation by a few." A book of which that can be said is most certainly an unusual book. Not only is Dr. Yerkes' statement probably true for different readers, but the same reader will undoubtedly go through all three of those stages in different parts of the book. In other words, it is a very uneven product. There are paragraphs that are valuable and suggestive. There are some that can be neglected and there are some that tend to induce rude handling. One questions whether Dr. Yerkes' well intentioned foreword is really a kindness to the author, for it leads the reader to expect more than he finds.

The author himself says in the preface, "The results of this work are here offered to physicians, social workers, and lay readers." In the opinion of your reviewer, this is too wide a field. One can not generally talk to an audience so widely diversified in interest and attainment. The difficulty of such an undertaking is well illustrated in the type of word the author has felt it necessary to define in his glossary. We find on the one hand, *acroparesthesia, angioneurotic Edema, fastidium, hebetude, mediastinal, pyosalpinx.* At the other end of the scale we have *aphasia, cardiac, cervical, clavicle, coryza, femur, olfactory, patella, psychosis, sinus infection, therapeutic, uterus.* Were it not for the statement already quoted, one would certainly be at a loss to guess for whom the author was writing.

The 200 cases are exceedingly varied in interest, in fullness, in the type of discussion, and in value. For example, case 182 is covered in eight words: "Female, Fourth Decade, Hyperthyroidism, no significant adjusted disfunction." The cases are the result of what the author calls a "survey" of the psychopathic cases in a town of 30,000 inhabitants in the Mississippi Valley.

It would seem that the term "survey" which the author uses repeatedly, is hardly appropriate for there is no indication that he made a systematic examination of the 30,000 inhabitants or of any unselected group of them. He spent a year in the town and saw such patients as came to him, or were sent to him by other physicians. So at best, he saw only a relatively few of those who were sufficiently bad off or had sufficient insight to conceive that they might be helped by a physician.

As to the method of treatment, there seems to be nothing new. It might be described as a modified psychoanalytic procedure. His attitude toward Freud and the analyst is perhaps fairly indicated by two quotations. "If I suspect that her inhibition of recall processes is too firmly fixed as a reactive habit to enable her to overcome it by relaxing and thereby acquiring a noninhibitive reactive set for purposes of examination I may be compelled to study her dreams, as interpretable indirect reactions; but one rarely encounters a patient who is unable to disclose *by ordinary processes of recall* ample material for an explanation

of her nevousness" (p. 207). And again "The family physician who is familiar with the general outline of Freud's teachings is in a far more favorable position to make effective his own methods of dealing with nervousness than is one who must confess that psychoanalysis is for him a *terra incognita*" (p. 208).

In Chapter V the author states "The special foundations of psychopathology are neural morphology, neural physiology, endocrinology, and behaviorism" and again "This book is exclusively concerned with the kinds of nervousness that are not due to any known infection, intoxications or structural changes of the nervous system, that do not require institutional management and that the internist ought, with few exceptions, to accept as coming within his general field of therapeutic and diagnostic endeavor." This chapter is a short but very clear statement of the foundations of psychopathology.

Chapter VI is again rather elementary, devoting itself to an explanation of the neurons, the cell body, axones, etc.

Chapter VII sets forth among other topics, "*The Hamilton Method of Investigating Reactions to Baffling Disadvantages,*" and is based upon experimental studies of 31 human subjects, 10 monkeys and baboons, 16 dogs, 5 cats, 1 horse, 1 mouse, 5 gray rats, 4 black rats, 8 albino rats and 6 gophers.

The remaining chapters contain little that is new to the student of psychology and psychoanalysis, though there are many interesting paragraphs and striking sentences. Yet they suffer as does the whole book, in the opinion of the reviewer, from the unfortunate attempt of the author to write one book for both the laymen and the professional.

Although the book probably adds nothing to our knowledge of psychopathology, there is set forth a method based upon Freudianism (but without its mystical interpretations)which will undoubtedly prove of value to students in a certain stage of their development.

H. H. GODDARD,
Ohio State University.

Helmholtz's Treatise on Physiological Optics. Translated from the Third German Edition. Edited by James P. C. Southall. Vol. III. Published by the Optical Society of America, 1925. Pp. x, 736 and 6 plates.

The third volume of the third German edition of the *Optik* appeared in 1910, one year before the second volume. The editor of this volume, von Kries, followed the text of the first edition and included extensive notes of his own. The third volume of the English translation follows the text of the third German edition and includes, in addition, two short notes by von Kries, a partial bibliography of some 550 articles which have appeared on vision between 1911–1925, an index of subjects, an in-

dex of authors, *corrigenda* in volume II, and additional *corrigenda* in volume I.

This third volume, on "The Theory of the Perceptions of Vision," of Helmholtz's monumental work will be of even greater service to psychologists than the second volume. Here one finds Helmholtz treating of perception in general, monocular and binocular vision, perception of depth, direction and rivalry, and theories; and appendices by von Kries on spatial configuration and questions of innate dispositions and experience. The work stands as a great and unrivalled classic for the foundation of experimental and physiological psychology on the side both of fact and method. The science has still to wait for a work which will begin to do for complex mental processes what Helmholtz's genius and labor did for sensation and perception.

What has already been said about the excellence of the preparation of the first two volumes is equally true of the third. Great credit must be given to Professor Southall and his collaborators for the successful manner in which they have accomplished a difficult and formidable task. Professor Southall wishes that tribute be paid also to Mr. Adolph Lomb whose "noble and singleminded devotion to the advancement of science for the welfare of mankind is so genuine and unstinted that I know he will consider himself amply rewarded if the usefulness of Hemholtz's work is continued and extended by this English edition of the Treatise on Physiological Optics."

C. C. PRATT,
Harvard University.

SEBA ELDRIDGE. *Political Action: a Naturalistic Interpretation of the Labor Movement in Relation to the State.* (Lippincott's Sociological Series.) Philadelphia, J. B. Lippincott Company, 1924. Pp. xviii, 382.

The author states that the work is an attempt to use the methods of investigation, with reference to certain political and economic problems, which have been applied in the fields of education and criminology, for the enrichment of the social sciences. These problems include (1) the relations of the capitalist and the wage-earning classes in the United States, (2) the future of representative government, (3) the struggle of social classes for power, and (4) the use of "direct methods" in the furtherance of group interests.

The method followed throughout has been the analysis of specific situations to determine causal relations of various factors to given situations, which are productive of change. The book presents a statement of the obstacles that oppose the successful operation of democracy. The author dwells on habit, instincts, and the power of the "suggestioning agencies" used by the dominant class as the determining fac-

tors in social problems. When the problem of class struggle, for example, is analyzed, it is shown that natural factors having nothing in common with reason produce social conditions. When the hereditary mental equipment evolved under primitive conditions is required to function under conditions of civilized life, the results are unsatisfactory. Neither faction in the class struggle is moved to action by reason and judgment, but each responds to instinctive impulses. No permanent equilibrium between the two forces, therefore, appears possible. Since one faction increases its strength at the expense of the other, "the outcome of the present situation will probably be the complete ascendancy of the one class or the other." The editor in his introduction well concludes "this book is calculated to lead either to a pessimistic conclusion about the future development of democratic society, or. . . . to vigorous efforts to set free those psychic processes which are the essence of democratic action."

The reader is prepared for the presentation of the results of the investigation not only by the editor who points out debatable positions assumed by the author but also by the author who recognizes the proponents of the influence of hereditary mental characteristics and of the results of custom, traditions, institutions, educational discipline, and other environmental factors. The final chapter is devoted to a discussion of these divergent views.

The book will contribute to any reader who desires to adapt himself to his environment and who wishes to learn about himself and his neighbors. It points out a more accurate way of estimating the capacities of people to function in any situation.

EDWIN B. SMITH,
Ohio University.

HANBURY HANKIN. *Common Sense and Its Cultivation.* E. P. Dutton, N. Y., Kegan Paul, London, 1926. viii and 289 pp. $2.50.

The author, a British physician in India, raises and attempts to answer two questions: Can we know more about the nature of common sense? Can its development be "stimulated by any practical changes in our educational methods?"

The book really deals with various problems which are usually grouped under the category of "the subconscious." For Dr. Hankin that is what common sense finally reduces to. So the thesis is presented that all of the important problems in commerce and finance, medicine and education can be solved more quickly and safely by the snap-judgments of the "subconscious" and "preconscious" minds than by the ratiocinations of the third "mind"—the conscious one.

For some reason or other, the author seems quite out of patience with education, particularly with the education of business men, for

whom the volume appears to have been written. We read in the preface that "our educational methods, whatever their matter, are fundamentally vicious in their manner" and that the study of foreign languages (p. 209) "is likely to have the disadvantage that his (the student's) capacity for developing the business instinct will be impaired."

Neither of Dr. Hankin's objections—to the artificiality and sinister influence of education and the omniscience of what is called "the subconscious" is new. Rousseau held much the same general point of view, and Hankin's "common sense" looks for all the world like the same thing as Bergson's "faculty of intuition."

The thesis of the volume is supported principally by weak evidence of the anecdotal sort and by unadulterated *ad hoc* logic. For example, we read (p. 54) that "The speed of lightning calculators in producing their results may be compared to the speed with which we arrive at common sense decisions. If we can explain this speed in one case, we may hope to have some clue to its nature in the other."

Prof. C. S. Meyers, who contributed the introduction, says that, "Although as Dr. Hankin admits, the book cannot always claim psychological precision . . . . it will make the reader think." Psychologically the book performs the office of emphasizing, from the point of view of a shrewd "lay" observer, the importance of pushing the *degrading* process in "learning" as near the asymptote as possible where this "learning" is of certain professional or vocational types. Otherwise it is trivial.

SAMUEL RENSHAW,
Ohio State University.

LEONARD DARWIN. *The Need for Eugenic Reform.* D. Appleton & Company, New York, 1926. 529 pp.

This book may seem at first sight to have a large amount of commonplace material needlessly written, but not so. The author in the introduction says it was not prepared for those trained in biology but rather for the well-educated person without special training who is interested in eugenics.

He begins by positing as his first principle that evolution always means change but not necessarily for the better. Hence, society needs eugenics to direct the process of human evolution. The Galtonian and Mendelian principles of inheritance are explained and applied to the transmitting of mental qualities.

Here the author gives a full and intelligent presentation of the problems of environment, the work of reformers, and how increasing populations have brought new perplexities. It is pointed out that practically all attempts at social reform are transient while eugenic reform would be permanent because it changes the hereditary factor.

Racial poisons such as drugs, alcohol, venereal diseases and their consequent effect upon racial deterioration are dealt with in a plain way. It is not meant to be exhaustive but rather to give the average reader a true picture of actual conditions.

A commonsense view of the Lamarchian theory of acquired characteristics is encouraged. It is shown that natural selection in society will not result in improvement, and that scientific methods of elimination of the unfit must be taken up. The subjects of segregation, sterilization and birth control are discussed as to their effectiveness and merits.

A long discussion of the modern social problems of feebleminded-ness, insanity, epilepsy, and tuberculosis follows, with careful statistics. Facts are brought to sustain the theory that the more intelligent classes are dying out and race deterioration is setting in. As a means of preventing this trend the author emphasies the help that eugenics can give. He advocates redistribution of wealth so that the strong middle stock families may be of average size and have ample education. Birth control and sex education are encouraged.

The subject of divorce is taken up with an aim of finding the cause. More careful supervision of marriage is advocated to prevent unfit from marrying. In short the whole book is an emphasis upon public education and bringing the facts of hygienic living before society.

As a conclusion Darwin discusses the purposes of life and the riddle of the universe. He concludes with a belief in scientific determinism and that while the universe may be teleological yet man must mold his own destiny. Here is where religion and science must meet.

The book is a well-written but somewhat wordy presentation of the needs of society, prepared in a way that is suited to the intelligence of the average mind.

STANLEY F. ROUSH,
Ohio University.

NEW BOOKS AND PAMPHLETS RECEIVED*

Books and pamphlets for review should be sent to James P. Porter, Department of Psychology, Ohio University, Athens, Ohio.

An Approach to Social Medicine. FRANCIS LEE DUNHAM. The Williams and Wilkins Company, Baltimore, Md., 1925. Price $4.00. 242 pp.

The Borderland: Some of the Problems of Insanity. THEO. B. HYSLOP. George H. Doran Company, New York. Price $3.00. 300 pp.

The Child and Religion. HENRY JONES, C. F. G. MASTERMAN, GEORGE T. LADD, AND OTHERS. G. P. Putnam's Sons, New York, 1905. 371 pp.

Employee Representation. ERNEST RICHMOND BURTON. The Williams & Wilkins Company, Baltimore, Md., 1926. Price $3.00. 283 pp.

The Great Abnormals. THEO. B. HYSLOP. George H. Doran Company, New York. Price $3.50. 289 pp.

G. Stanley Hall: A Biography of a Mind. LORINE PRUETTE. D. Appleton & Company, New York, 1926. Price $2.50. 267 pp.

The Hereditary Factor in the Etiology of Tuberculosis. ALBERT GOVAERTS. Eugenics Record Office, Bulletin No. 23. Cold Spring Harbor, Long Island, N. Y. 20 pp.

The History of Utopian Thought. JOYCE ORAMEL HERTZLER. The Macmillan Company, New York City, 1926.

The Inheritance of Mental Diseases. ABRAHAM MYERSON. The Williams & Wilkins Company, Baltimore. Price $5.00. 336 pp.

Introduction to Retail Advertising. ARTHUR JUDSON BREWSTER. A. W. Shaw Company, New York, 1926. 319 pp.

Life of Christ. RUFUS M. JONES. American Library Association, Chicago, 1926. Price 50 cents. 28 pp.

Labor Relations in Industry. DWIGHT LOWELL HOOPINGARNER. A. W. Shaw Company, Chicago, Ill. Price $4.00. 553 pp.

Methods of Handling Test Scores. LUELLA COLE PRESSEY AND SIDNEY L. PRESSEY. World Book Company, 1926. 60 pp.

Mongrel Virginians. ARTHUR H. ESTABROOK AND IVAN E. MCDOUGLE. The Williams & Wilkins Company, Baltimore, Md., 1926. Price $3.00. 205 pp.

---

*Mention here does not preclude further comment.

The Need for Eugenic Reform. MAJOR LEONARD DARWIN. D. Appleton & Company, New York, 1926. 529 pp.

Our Children. M. V. O'SHEA. American Library Association, Chicago, Ill., 1925. 34 pp.

Poor Relief in Pennsylvania. EMIL FRANKEL. Department of Welfare Bulletin No. 21, 1925. Harrisburg, Pa. 149 pp.

A Practical Psychology of Babyhood. JESSIE C. FENTON. Houghton Mifflin Company, Boston, Mass. 348 pp.

The Psychology of Handwriting. ROBERT SAUDEK. George H. Doran, Company, New York. Price $5.00. 288 pp.

Psychologies of 1925. MADISON BENTLEY, KNIGHT DUNLAP AND OTHERS. Clark University, Worcester, Mass., 1926. 412 pp.

Psychology and Its Uses. EVERETT DEAN MARTIN. American Library Association, Chicago, 1926. 47 pp.

Psychology and Education. ROBERT MORRIS OGDEN. Harcourt, Brace & Co. New York City, 1926.

The Psychology of Gestalt. HARRY HELSON. University of Illinois, Urbana, Ill. American Journal of Psychology. 229 pp.

The Retarded Child: How to Help Him. ARNOLD GESELL. Public School Publishing Company, Bloomington, Ill., 1925. 100 pp.

Religion in Everyday Life. WILFRED T. GRENFELL. American Library Association, Chicago, 1926. Price 50 cents. 30 pp.

Scholastic Aptitude Tests. CARL C. BRIGHAM, Chairman of Committee of Review, Princeton University. A manual for the use of college officers as a college entrance examination.

School as the People's Clubhouse, The. HAROLD O. BERG. Physical Education Series No. 6, 1925. Department of the Interior, Washington, D. C. 16 pp.

Socializing Education through Community Welfare Clubs. WALTER SCOTT MCNUTT. The Palmer Company, 120 Boylston Street, Boston, Mass. Price $1.00. 54 pp.

Summary of Investigations Relating to Reading. WILLIAM SCOTT GRAY. University of Chicago Press. Chicago, Ill., 1925. 275 pp.

Survey of The Winnetka Public Schools. CARLETON WASHBURNE, MABEL VOGEL AND WILLIAM S. GRAY. Public School Publishing Co. Bloomington, Ill., 1926. 135 pp.

Standard Service Arithmetics, Book Two. F. B. KNIGHT, J. W. STUDEBAKER, AND G. M. RUCH. Scott, Foresman & Co., New York.

Temperament and Race. S. D. PORTEUS AND MARJORIE E. BABCOCK. Richard G. Badger, Boston, Mass. Price $3.00. 364 pp.

A Universal Scale of Individual Performance Tests, Examination Manual. PAUL C. SQUIRES. Princeton University Press, Princeton, N. J., 1926. 158 pp.

A LIST OF CONTRIBUTIONS TO EDUCATION, TEACHERS COLLEGE, COLUMBIA UNIVERSITY, NEW YORK CITY

The Ability of High School Pupils to Select Essential Data in Solving Problems. BENJAMIN W. DAILY. No. 190.

Biology in Secondary Schools and the Training of Biology Teachers. CHARLES WILLIAM FINLEY. No. 199.

Certain Phases of the Administration of High School Chemistry. WESTON WILLIAM CARPENTER. No. 191.

Children's Social Concepts. A Study of Their Nature and Development. HYMAN MELTZER. No. 192.

A Detailed Analysis of Achievement in the High Schools. CECIL WHITE FLEMMING. No. 196.

Educational Attitudes and Policies of Organized Labor in the United States. PHILIP R. V. CUROE. No. 201.

Effect of Age and Experience on Tests of Intelligence. VERNON A. JONES. No. 203.

Graded Units in Student-Teaching. HUGH CLARK PRYOR. No. 202.

A Method of Directing Children's Study of Geography. EDWIN HEWETT REEDER. No. 193.

A Study of the Group Method of Measurement of Sight-singing. RAYMOND M. MOSHER. No. 194.

The Supervision of Student-Teachers in Religious Education. FAYE HUNTINGTON KLYVER. No. 198.